Selection of Individual $V_H$ Genes Occurs at the Pro-B to Pre-B Cell Transition

Wenzhao Meng, Lenka Yunk, Li-San Wang, Avinash Maganty, Emily Xue, Philip L. Cohen, Robert A. Eisenberg, Martin G. Weigert, Stephane J. C. Mancini and Eline T. Luning Prak

*J Immunol* 2011; 187:1835-1844; Prepublished online 11 July 2011; doi: 10.4049/jimmunol.1100207

http://www.jimmunol.org/content/187/4/1835

Supplementary Material

http://www.jimmunol.org/content/suppl/2011/07/11/jimmunol.1100207.DC1

References

This article cites 70 articles, 33 of which you can access for free at:
http://www.jimmunol.org/content/187/4/1835.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

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2011 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Selection of Individual \( V_H \) Genes Occurs at the Pro-B to Pre-B Cell Transition

Wenzhao Meng,* Lenka Yunk,* Li-San Wang,* Avinash Maganty,* Emily Xue,* Philip L. Cohen, ‡ Robert A. Eisenberg, ‡ Martin G. Weigert, ‡*,‡,* Stephane J. C. Mancini, ‡ and Eline T. Luning Prak*

B cells are subjected to selection at multiple checkpoints during their development. The selection of Ab H chains is difficult to study because of the large diversity of the CDR3. To study the selection of individual Ab H chain V region genes (\( V_H \), we performed CDR3 spectratyping of ~75–300 rearrangements per individual \( V_H \) in C57BL/6J mice. We measured the fraction of rearrangements that were in-frame in B cell DNA. We demonstrate that individual \( V_H \)s have different fractions of in-frame rearrangements (IF fractions) ranging from 10 to 90% and that these IF fractions are reproducible in different mice. For most \( V_H \)s, the IF fraction in pro-B cells approximated 33% and then shifted to the nearly final (mature) B cell value by the cycling pre-B cell stage. The frequency of high in-frame (IF) \( V_H \) usage increased in cycling pre-B cells compared with that in pro-B cells, whereas this did not occur for low IF \( V_H \)s. The IF fraction did not shift as much in BCR-expressing B cells and was minimally affected by L chain usage for most \( V_H \)s. High IF clan II/III \( V_H \)s share more positively charged CDR2 sequences, whereas high IF clan I J558 CDR2 sequences are diverse. These data indicate that individual \( V_H \)s are subjected to differential selection, that \( V_H \) IF fraction is mainly established through pre-BCR-mediated selection, that it may operate differently in clan I versus II/III \( V_H \s), and that it has a lasting influence on the Ab repertoire. *The Journal of Immunology, 2011, 187: 1835–1844.

Immunglobulin loci of developing B cells undergo a series of DNA rearrangements [V(D)J recombination] that culminate in the assembly of Ab H and L chain variable regions (1, 2). A large and diverse repertoire of Abs is created by the recombination of multiple V, D, and J segments, variability at the junctions between these gene segments, and numerous combinatorial possibilities for H plus L chain pairing. But this diversity comes at a cost: nonproductive (out-of-frame; OF) rearrangements occur frequently and, among the rearrangements that are productive (in-frame; IF), many of the resulting Abs are autoreactive (3).

The manner in which the primary Ab repertoire is purged of autoreactive receptors is fundamental to the understanding of self-tolerance. It is typically assumed that the major stage of bone marrow (BM) B cell development where censoring of the primary Ab repertoire takes place is at the pre-B to naive B cell transition. At this stage, B cells with autoactive IgM Abs can edit their Ab receptor specificity by undergoing further L chain rearrangement (reviewed in Ref. 4) or undergoing clonal deletion. It seems logical for editing of autoactivity to occur after the full Ab (H plus L chain) has been formed, but several lines of evidence suggest that Ab H chains are also subjected to specificity-based selection during early B cell development.

For example, Ab H chains are thought to undergo “structural selection” at the pro-B to pre-B cell transition. Consistent with this model, H chains that pair well with the surrogate L chain result in a pre-BCR that is capable of signaling via associated Igα and Igβ domains (5), resulting in downregulation of the V(D)J recombinase and IL-7–dependent proliferation in large, BM cycling pre-B cells [Fr. C’; Hardy BM fraction C’ (6–8)]. Ten Boekel et al. (9) showed that approximately half of Ab H chains were able to pair well with the surrogate L chain and promote B cell development. But why some H chains pair well with surrogate L chain and others pair poorly is incompletely understood, in part because early experiments did not distinguish between selection based on CDR3 (which generally derives a minority of its sequence from the \( V_H \) segment) versus selection elsewhere in the \( V_H \). Furthermore, the mechanism of pre-BCR signaling is unclear (e.g., is it ligand dependent?). Compelling data indicate that galectin 1, secreted by BM stromal cells, binds to the \( V_\alpha \) unique region of the pre-BCR and influences pre-B cell differentiation and proliferation (10, 11). In contrast, the crystal structure of the human pre-BCR Fab, along with other biochemical data, suggest that the long tails of the surrogate L chains could mediate the self-assembly of multimers (5, 12, 13). Furthermore, recent data implicate a role for N-linked glycosylation within the H chain C region in surrogate L chain binding and pre-BCR cross-linking (14).

Other recent evidence suggests that Ab H chains in developing pre-B cells are subjected to distinctive selection, particularly with regard to autoreactivity: surrogate L chain knockout mice develop...
spontaneous autoimmunity, including the production of antinuclear Abs (15). The H chains of Abs from surrogate L chain knockout mice are enriched for positively charged arginine residues in the CDR3, resembling the canonical sequences of anti-DNA Abs (15, 16). In the absence of Ψ, these H chains were expressed on the cell surface, bound to nuclear Ags such as DNA, and permitted B cell proliferation. In contrast, in the presence of surrogate L chain, freshly isolated anti-DNA pro-B cells were reduced in frequency after a few days in culture, suggestive of negative selection. Additionally, two groups have demonstrated that H chains with CDR3s that mimic or are derived from known autoantibodies promote surrogate L chain-independent B cell maturation and proliferation (17, 18).

Work from other laboratories supports the selection of D$_H$ reading frames, CDR3 length, and amino acid content during early B cell development (3, 9, 18–23). Earlier studies of Ab H chain selection also revealed a progressive shift in the frequency of V$_H$ usage, particularly within the 71B3 family, from 3’ (D-proximal) in the earliest stage B cells to more 5’ V$_H$ usage in more mature B cell subsets. This shift in V$_H$ usage is recapitulated during ontogeny, with fetal B cell repertoire exhibiting increased 3’ V$_H$ usage (24–27). However, the analysis of selection based on V$_H$ frequency is problematic because multiple factors, including the frequency of rearrangement, can independently influence the observed V$_H$ usage (28). Furthermore, the frequency of V$_H$ rearrangement is not necessarily proportional to whether the V$_H$ is positively or negatively selected. A case in point is the most 3’ V$_H$, 81X, which is frequently but often nonproductively rearranged (29). The analysis of CDR3 sequences of certain V$_H$s, including 81X, VH12, and Q52 (9, 22, 23, 30, 31), clearly documents specific changes in the repertoire that occur during the pre-BCR selection step. These and other studies have led to the prevailing view that the CDR3, which contributes the lion’s share of H chain diversity (32), is the critical region for H chain selection (e.g., Refs. 33, 34). In point of fact, however, the selective consequences of V$_H$ sequences outside the CDR3 are almost entirely unknown. Only a handful of studies, focused on specific V$_H$s such as 81X, VH11, and VH12, have documented selection of amino acids outside of the CDR3 (23, 30, 35).

A more systematic study of V$_H$ selection would enhance our understanding of tolerance mechanisms that apply specifically to Ab H chains. For example, V$_H$ replacement is a mechanism of Ab H chain editing that exchanges a V$_H$ in an existing V(D)J rearrangement with a new VH by invasion into a cryptic heptamer near the 3’ end of the V$_H$ (36–38). As such, V$_H$ replacement does not remove the existing CDR3, but rather adds to it and swaps out the V$_H$. If the CDR3 is the dominant factor in H chain selection, this receptor editing mechanism should be ineffective. Understanding the developmental timing of V$_H$ selection would also provide better insights into how the pre-BCR versus the BCR are involved in autoantibody regulation (15).

Accordingly, we have applied a more general approach to the issue of H chain selection by measuring the fractions of in-frame rearrangements (IF fractions) of 20 different V$_H$s in genomic DNA of B cells from C57BL/6J (B6) mice. In the B6 spleen, we found that different V$_H$s had different IF fractions, ranging from 10 to 90%. Most of the clan I V$_H$s, including all but one member of the J588 family that were evaluated, had a high IF fraction. In a more detailed analysis of BM B cell subsets, most V$_H$s exhibited an IF fraction of ~33% at the pro-B cell stage, and the IF fraction increased substantially for high IF V$_H$s by the large cycling pre-B cell stage. These findings suggest that pre-BCR–driven selection influences the V$_H$ IF fraction. In support of this hypothesis, the IF fraction appears to be positively correlated with an increase in rearrangement frequency as cells progress from the pro-B to the pre-B stage. In contrast to the large changes in the V$_H$ IF fraction at the pro-B to pre-B cell transition, changes in BCR-expressing B cell subsets tended to be more subtle, and, for most H chains, L chain usage did not significantly alter the IF fraction. Taken together, these findings suggest that the major shift in the IF fraction occurs at the pro-B to pre-B cell transition and involves pre-BCR rather than BCR-based selection.

**Materials and Methods**

**Mice**

B6 mice were purchased from The Jackson Laboratory (Bar Harbor, ME) and maintained at the University of Pennsylvania School of Medicine under an institutional animal care and use committee-approved protocol. Vo8 L chain knock-in mice were maintained at the University of Chicago under an institutional animal care and use committee-approved protocol.

**Flow cytometry**

Cell suspensions from 3- to 4-mo-old B6 mice were prepared from femurs, tibias, and spleens in FACS buffer (PBS, 0.5% BSA, 0.01% NaN$_3$, 1 mM EDTA) after RBC lysis (ACK Lysing Buffer; BioWhittaker, Walkersville, MD). All Abs were purchased from eBioscience (San Diego, CA), BD Biosciences (San Jose, CA), and BioLegend (San Diego, CA). The following Ab–fluorophore combinations were used to resolve BM subsets: FITC anti–heat-stable Ag (30F1), PE anti–CD43 (57), PE–Cy7 anti-IgM (II/41), PE–Cy5.5 anti–CD19 (1D3), allophycocyanin anti–CD93 (AA4.1), allophycocyanin–AF750 anti–B220 (RA3-6B2), and Pacific blue anti–IgD (11–26). BM B cell fractions are defined as follows based on the scheme by Hardy et al. (7): Fr. B–C (B220+CD19+CD43+IgM+CD24$^+$/CD93$^-$), Fr. C’ (B220+CD19+CD43+IgM+CD24$^+$/CD93$^+$), Fr. D (B220$^+$ CD43+IgM+CD24$^+$/CD93$^-$), Fr. E (B220+CD19+CD43$^+$IgM+CD24$^+$/CD93$^+$), and Fr. F (B220+CD19+CD43$^+$IgM$^-$CD24$^+$/CD93$^+$). The following Abs were used to resolve splenic B cell subsets: FITC anti–IgM (II/41), PE anti–CD21 (7G6), allophycocyanin anti–CD93 (AA4.1), and allophycocyanin–AF750 anti–B220 (RA3-6B2). Splenic B cell subsets were defined as follows: transitional (Trans; B220$^+$CD93$^-$), follicular (FO; B220$^+$CD93$^+$IgM$^-$CD21$^+$), and marginal zone (MZ; B220$^+$CD93$^+$IgM$^+$CD21$^-$). For all sorts, dead cells were first eliminated by DAPI staining and cell doublets/aggregates by pulse width gating. The purities of the Fr. B-C sorted cell populations were 98–99%. FACs analysis and sorting were performed on the LSR II and FACSAria cytometers, respectively (BD Biosciences), at the University of Pennsylvania Flow Cytometry core facility. Flow cytometry data were analyzed using FlowJo software (version 7.5.5; Tree Star, Ashland, OR). For sorting of λ$^+$ B cells, splenic B cells were stained with allophycocyanin anti–λ (RML–42), FITC anti–κ (187.1), allophycocyanin–AF750 anti–B220, and PE–Cy5.5 anti–CD19.

**CDR3 spectratyping**

CDR3 spectratyping was performed to evaluate the V$_H$ IF fractions. Genomic DNA was isolated from either spleens or sorted cells, purified according to the manufacturer’s directions using PureGene (Qiagen, Valencia, CA), and amplified using the V$_H$ primers in Supplemental Table I and a labeled reverse primer in J$_H$(5’-FAM-CTG TGA GAG TGG TGC GCT G-3’). All primers were synthesized by Integrated DNA Technologies (Coraville, IN). Each 20-μl PCR contained 5–50 ng genomic DNA, 1× PCR buffer (AmpliTaq Gold 10× PCR buffer with 15 mM MgCl$_2$; Roche Applied Sciences, Indianapolis, IN), 0.2 mM of each dNTP (dATP, dCTP, dGTP, and dCTP; Promega, Madison, WI), 0.6 μM of each primer, and 1 U of DNA polymerase (AmpliTaq Gold; Roche Applied Sciences). The cycling conditions were 10 min at 94°C, followed by 40 cycles of 94°C for 30 s, T$_m$S (Supplemental Table I) for 30 s, and 72°C for 30 s, followed by 20 min at 72°C. To minimize notching artifacts due to nonuniform adenylate addition by the Taq polymerase, some V$_H$ primers were modified by “PIGtailing” (incorporating the sequence GTTTCTT at the 5’ end; see the italicized sequences in Supplemental Table I). This modification resulted in nearly 100% adenylation of the 3’ end of the FAM-labeled opposite strand, consistent with what has been described previously (39). Two microliters of PCR products was resolved by capillary electrophoresis on an ABI 3100 analyzer (Applied Biosystems, Foster City, CA).

**CDR3 length analysis**

Capillary electropherograms were generated and analyzed using ABI Genetyper 3.7 or ABI Peak Scanner Software, version 1.0. Peak sizes were interpolated using a ROX ladder “DS30” (Applied Biosystems) alone or...
combined with our homemade ladder (sizes of the homemade ladder are 195, 198, 201, 204, 207, 210, 213, 216, 219 bp). The IF fraction was calculated by determining the number of peaks that are IF and dividing by the total number of peaks (IF + OF). The CDR3 lengths were calculated, beginning with the conserved cysteine residue at the 3’ end of the VJ (position 96) and ending with the last conserved glycine in the middle of JH2.

**CDR3 peak count metadata**

In Fig. 2B, the number of peaks and IF fraction (the latter in percent) for each VJ are as follows: J606.1, Fr. B-C n = 109, 32%; VH10-VH12 n = 55, 42%; VH15 n = 84, 29%; VH10 n = 91, 69%; 3609 n = 109, 42%; J606.1 n = 178, 81%; VH12 n = 218, 21%; 3609N.2 n = 371, 36%; SI107.3/4 n = 107, 82%; VJ11 n = 109, 27%; SM1.7/24 n = 77, 88%; SI107.1 n = 78, 101%; Q52.8/5 n = 63, 38%; 7183.9 n = 68, 75%; 7183.4 n = 100, 60%; Q52.2 n = 86, 45%; 7183.2 n = 93, 13%. In Fig. 3A, the number of peaks and IF fraction analyzed for each subset are as follows: J606.1, Fr. B-C n = 83, 39%; Fr. C’ n = 231, 87%; Fr. 10, Fr. B-C n = 84, 30%; Fr. C’ n = 69, 79%; VH558.47, Fr. B-C n = 56, 38%; Fr. C’ n = 22, 64%; VH3609N.2, Fr. B-C n = 48, 35%; Fr. C’ n = 31, 35%; VH7183.2, Fr. B-C n = 69, 32%; Fr. C’ n = 55, 29%. In Fig. 3B, the number of peaks analyzed for each subset is as follows: J606.1, Fr. B-C n = 83; Fr. C’ n = 231; VH558.85, Fr. B-C n = 42; Fr. C’ n = 88; VH3609.2, Fr. B-C n = 47; Fr. C’ n = 21; VH558.47, Fr. B-C n = 18; C’ n = 106, 69 Fr. C’ n = 124, 13%; VH15.47 n = 48, 35%; VH10 n = 83; VH7183.9 n = 56, 38%; Fr. C’ n = 48, 35%; VH10 n = 84, 30%; Fr. C’ n = 54, 54%; Trans n = 29, 25%; Fr. F n = 49, 39%; Fr. D n = 77, 91%; J558.85 n = 92, 46%; Fr. E n = 129, 82%; MZ n = 55, 42%; 3609N.2, Fr. B-C n = 48, 35%; Fr. C’ n = 31, 35%; Fr. D n = 92, 46%; Fr. E n = 76, 39%; Fr. F n = 61, 46%; Trans n = 47, 33%; FO n = 68, 31%; MZ n = 61, 38%; VH10 n = 84, 30%; Fr. C’ n = 69, 57%; Fr. D n = 53, 59%; Fr. E n = 50, 56%; Fr. F n = 54, 54%; Trans n = 29, 26%; FO n = 19, 68%; MZ n = 75, 68%. In Fig. 4B, the numbers of rearrangements and their IF fractions are as follows: 7183.2–VH12 n = 64, 14%; VH15–VH8 n = 61, 25%; 3609–VH28 n = 129, 25%; J558–VH28 n = 64, 39%; 3609–VH8 n = 109, 32%; VH10–VH8 n = 35, 69%; 3609–VH28 n = 49, 84%; 3609–VH11 n = 78, 84%; 3609–VH28 n = 100, 77%; J558–VH28 n = 77, 82%; SM1.7/24–VH8 n = 50, 90%; 7183–VH28 n = 10, 13%; VH12–VH8 n = 27, 15%; 3609–VH28 n = 29, 28%; VH10–VH8 n = 98, 82%; 3609–VH11 n = 55, 89%; J558–VH11 n = 31, 77%; 3609–VH8 n = 77, 91%; J558–VH28 n = 79, 85%; SM1.7/24–VH8 n = 17, 94%. For the VHs in this comparative analysis, the p values < 0.05 were considered statistically significant.

**Results**

Using CDR3 spectratyping, it is possible to survey the diversity of a single set of VH and JH gene rearrangements that include any of the DH genes (Fig. 1A). When many Ab gene rearrangements are amplified from a peripheral B cell population, the CDR3 spectratype typically assumes a Gaussian distribution with peaks that appear at three nucleotide intervals because of the overrepresentation of IF rearrangements. But when single VH–JH primer sets are used at limiting amounts of input DNA, it is evident that some of the rearrangements are OF (Fig. 1B, asterisks). To define a rearrangement as having an IF length, the number of nucleotides in the CDR3 that spans the 3’ end of the VJ through the D gene segment and into the 5’ end of the JH gene segment must be an integral multiple of three (Fig. 1C). All rearrangements that fail this test of size are, by definition, OF. However, not all sequences that are of the correct IF length are productive because they could harbor stop codons. Thus, the measured IF fraction, [IF/(IF + OF)] × 100%, is a maximal estimate of the true IF fraction.

To obtain an accurate measurement of the frequencies of individual rearrangements by CDR3 spectratyping, B cell genomic DNA was diluted to very low concentrations so that it was unlikely that more than one rearrangement corresponded with a peak of a given size. The strategy was to perform several independent amplifications of the same DNA sample and count the numbers of IF versus OF sized peaks to determine the overall IF fraction. To prove that single PCR amplifications were detecting only one rearrangement per peak, amplicons of a given size were cloned and sequenced from single reactions. In each case, only one sequence of a given size was repeatedly recovered per reaction, as expected if each peak in a single reaction corresponds with one and only one rearrangement (Supplemental Fig. 2). This problem is analogous to estimating the likelihood of two people in a group having the same birthday, and the “collision probability” can be modeled using different underlying distributions of rearrangement data. Assuming a normal distribution with a SD of 10 peaks, a range of 60 rearrangement lengths, and 3 peaks per PCR (which resemble our typical reaction conditions), the collision probability is ~10%.

The IF fraction differs in different VH through the H chain locus

If VH are not subjected to selective pressures, the expected IF fraction is ~70% (Fig. 1D). This assumes a simple model with instant and complete feedback inhibition of rearrangement once an IF rearrangement is produced and a maximum of two rearrangement attempts (27). Alternatively, if a particular VH were favored, one would expect more of the B cells to express that VH as an IF rearrangement than as an OF rearrangement (Fig. 1E). Conversely, if a VH were counterselected, more B cells should harbor that VH as an OF rearrangement than as an IF rearrangement (Fig. 1E).

To determine if different VHs have different IF fractions, we first analyzed the IF fractions of two VHs (J606.1 and 3609N.2) that are closely positioned in the middle of the germline H chain locus in B6 mice. The measured IF fractions of J606.1 and 3609N.2 rearrangements were 80 and 34%, respectively. These findings were reproducible in two individual B6 mice (Fig. 2A). Differences in the IF fraction can be modeled using binomial probabilities (see Materials and Methods), and the difference between J606.1 and 3609N.2 is highly significant (n = 200 rearrangements per group, p < 1 × 10−10). To validate these findings further, we performed DNA sequence analysis. Nine of 11 (~80%) sequences cloned from J606.1–JH2 and 3 of 8 (~40%) from 3609N–JH2 had IF VH rearrangements, consistent with the spectratyping results (data not shown).
To determine if other VHs also differ significantly in their IF fractions, we next examined the IF fractions among other VH families, including the most D-proximal VH (7183.2, equivalent to 81X in BALB/c), the most D-distal VH (J558.88), some B-1 lineage-associated VHs (VH11, VH12, Q52, S107), some that are overused in antinuclear Abs [J558.85 (VMU3.2), J558.72 (VH124), 7183.9], and one that is overused in non-antinuclear Abs [SM7 (V130) (44)]. [Note that the VH nomenclature for B6 is applied to all VH names (43) and DH names (45) in this article.]

The measured IF fractions of different VHs are displayed in Fig. 2B. At 75 peaks per VH, a difference of 10% or more in the IF fraction is statistically significant based on cumulative binomial probabilities (p ≤ 0.05). By this criterion, several of the VHs varied significantly in their IF fractions from one another, with

FIGURE 1. Measurement and theories of the VH IF fraction. A, Schematic of spectratyping PCR. Shown is a simplified H chain locus with two VHs (J606.1 and 7183.4), two DHs, two JH gene segments, and one CH. The primers (arrows) specifically amplify rearrangements that involve the J606.1 VH, either DJ or JH2. The reverse primer in JH2 has a fluorescent tag (star) that is used to analyze product sizes by capillary electrophoresis. Boxes denote exons, lines represent introns, and dashed lines show which gene segments participated in a VH-J606.1–DH–JH2 rearrangement. B, Spectratypes of VHJ606.1–JH2 rearrangements. Spleen DNA from 3-mo-old B6 mice was amplified using a VH-specific J606.1 primer and a FAM-labeled JH2 primer. PCR products from five individual reactions are displayed as spectratypes. In each spectratype, the y-axis represents the fluorescence intensity, and x-axis gives the product sizes in nucleotides (nt). OF peaks (OF) are marked with asterisks. C, Calculation of the VH IF fraction. The IF fraction of a single VH among a population of B cells can be envisioned as a ratio of the number of productive (VHDJH+)+ rearrangements to the total number of productive and non-productive rearrangements (VHDJH+ + VHDJH2). D, Model of VH IF fraction under conditions of uniform selection. According to a simplified model of H chain gene rearrangement, B cells undergo a maximum of two rearrangement attempts per cell, and rearrangement promptly ceases once a productive rearrangement has occurred (27). With estimated B cell genotype frequencies of 0.6 for VDJ+/DJ and 0.4 for VDJ+/VDJ2, the overall IF fraction (VDJ+/VDJ+ + VDJ+/VDJ2), according to this model, is (0.6 + 0.4)/(0.6 + 0.4 + 0.4) = 0.71. E, Model of VH IF fraction under conditions of differential VH selection. Under conditions of differential VH selection, VHs will have different IF fractions. According to this model, if a VH has a high IF fraction, then B cells that express a functional H chain rearrangement (darkly shaded cells) will be more frequent than B cells that harbor the same H chain as a nonfunctional rearrangement (lightly shaded cells). The converse occurs if a VH has a low IF fraction.

Average IF% = (0.6 + 0.4)/(0.6 + 0.4 + 0.4) × 100% = 71%
The major shift in the V\textsubscript{H} IF fraction occurs at the pre-B to pre-B cell transition

We considered two possible explanations for differences in the IF fraction among rearrangements of different V\textsubscript{H}\textsubscript{18}. The first was that successful rearrangement of different V\textsubscript{H}\textsubscript{18} occurred with different probabilities. The second possibility was that all V\textsubscript{H}\textsubscript{18} rearranged randomly (e.g., each started off with an IF fraction of 33%) and were subsequently subjected to differential selection. To distinguish between these alternatives, we analyzed the IF fraction at different stages of B cell development using sorted B cell subsets from the BM of B6 mice (Fig. 3A; see Materials and Methods for sort parameters). We analyzed the IF fraction of a high IF (J606.1), one intermediate IF (V\textsubscript{H}10), and three low IF (J558.47, 3609N.2, and 7183.2) V\textsubscript{H}\textsubscript{18} in BM pro-B cells. All five pro-B IF fractions are ~33%. Therefore, these data disfavor biased recombination as an explanation for the different V\textsubscript{H} IF fractions.

To evaluate the role of the pre-BCR in shaping the V\textsubscript{H} IF fraction, we next analyzed the IF fraction in cycling pre-B cells (Fr. C'). Consistent with earlier reports of H chain selection being

FIGURE 3. Establishment of the V\textsubscript{H} IF fraction occurs early during B cell development. A, V\textsubscript{H} IF analysis in BM pro-B cells (Fr. B-C) and BM cycling pre-B cells (Fr. C'). IF fractions (given in percent) for J606.1–JH2, J558.10–JH2, J558.47–JH2, 3609N.2–JH2, and 7183.2–JH2 rearrangements are plotted for each B cell subset. White bars denote Fr. B-C, and gray bars represent Fr. C'. Asterisks denote IF fractions that differ significantly between Fr. B-C and Fr. C': \( p < 0.05 \) (by binomial probability). B, V\textsubscript{H} IF peak number analysis in Fr. B-C and Fr. C'. Plotted are the frequencies of V\textsubscript{H}J606.1–JH2, J558.85–JH2, 3609N.2–JH2, J558.47–JH2, and 7183.2–JH2 rearrangements per 10,000 sorted B cells from Fr. B-C versus Fr. C'. The total raw peak numbers are given in Materials and Methods. Asterisks denote peak numbers that differ significantly between Fr. B-C and Fr. C': \( p < 0.05 \) (by \( \chi^2 \) test). Data for both panels in the figure are pooled from sorted cells of 3-mo-old B6 mice (n = 4–8 mice).

FIGURE 2. V\textsubscript{H}\textsubscript{18} differ in their IF fractions. A, Mouse-to-mouse reproducibility of the V\textsubscript{H} IF fraction. V\textsubscript{H} IF fraction analysis was performed on spleen DNA from 3-mo-old B6 mice (n = 2) for V\textsubscript{H} J606.1 (white bars) and 3609N.2 (gray bars). The numbers of J606.1–JH2 rearrangements and their corresponding IF fractions (the latter in percent) are as follows: mouse #1, 124 peaks, 80%; mouse #2, 194 peaks, 81%. For 3609N.2–JH2: mouse #1, 290 peaks, 36%; mouse #2, 293 peaks, 33%. The IF fractions calculated for individual mice are plotted for each V\textsubscript{H}. The difference in the IF fraction between the two V\textsubscript{H} is significant: \( p < < 1 \times 10^{-10} \) (by cumulative binomial probability). B, Survey of the IF fraction in 20 different V\textsubscript{H}\textsubscript{18}. V\textsubscript{H} IF fractions were measured in genomic DNA from 3-mo-old B6 mice spleen DNA (n = 2). V\textsubscript{H} IF fractions are arranged in the order of their location in the germline locus, with 5' V\textsubscript{H} on the left. The numbers of rearrangements and the IF fractions analyzed for each V\textsubscript{H} are given in Materials and Methods. Some of the primers bind to more than one V\textsubscript{H} (see Supplemental Table I for primer information). VHs are color-coded by family.

a range of IF fractions of 10–90% (Fig. 2B). Hereafter, we refer to the V\textsubscript{H} IF fraction as being “high” if it is >70%, “intermediate” if 50–70%, or “low” if <50% based on measurements in splenic B cells. Seventy percent is the cutoff predicted by the standard rearrangement model (Fig. 1D), whereas any IF fraction <50% requires some form of counterselection. [Even if all B cells underwent complete V(D)J rearrangement on both alleles, the minimum V\textsubscript{H} IF fraction in the absence of counterselection would be 50%.] To determine if the V\textsubscript{H} IF fraction is reproducible in different B6 mice, the data for each V\textsubscript{H} were analyzed separately for each mouse and compared between mice (for all V\textsubscript{H} having more than 20 peaks per mouse). This analysis revealed that the average difference in the IF fraction between different mice was 6%, which was not statistically significant (data not shown).

We analyzed the same data set with respect to CDR3 length (Supplemental Fig. 3A). To compare CDR3 lengths between different V\textsubscript{H}\textsubscript{18}, we defined the CDR3 as beginning with the conserved cysteine residue at the 3’ end of the V\textsubscript{H} (position 96) and ending with the last conserved glycine in the middle of JH2. This definition of the CDR3 length allowed for accurate comparison of V\textsubscript{H} sequence lengths when there is extensive nibbling of the JH. The following V\textsubscript{H} have additional nucleotides at the end of framework 3 (CAR): 3609.11 (4 nt), 3609.1 (6 nt), V\textsubscript{H}12 (6 nt), V\textsubscript{H}11 (2 nt), S107.1 (4 nt), Q52.8/5 (2 nt). Therefore, their mean CDR3 lengths are longer. There is no statistically significant correlation between the mean CDR3 length and the V\textsubscript{H} IF fraction if one stratifies the data based on V\textsubscript{H} (\( r^2 = 0.03 \)).
tied to pre-BCR expression (9), the IF fraction increased for the high and intermediate IF V<sub>HI</sub> (p < 0.05). Conversely, none of the low IF V<sub>HI</sub> exhibited a statistically significant increase in the IF fraction at the cycling pre-B cell stage.

The frequency of high IF VH rearrangements increases in cycling pre-B cells compared with that in pro-B cells

To determine if the V<sub>HI</sub> IF fraction is linked to expansion at the cycling pre-B cell stage, we compared the frequency of V<sub>HI</sub> rearrangements in Fr. B-C versus Fr. C’ (Fig. 3B). If the V<sub>HI</sub> IF fraction is linked to proliferation in Fr. C’, then B cells expressing a low IF V<sub>HI</sub> would be expected to undergo fewer rounds of cell division than B cells expressing a high IF V<sub>HI</sub>. Consistent with this prediction, the numbers of J606.1 and J558.85 (both high IF V<sub>HI</sub>) rearrangements increased in Fr. C’ compared with those in Fr. B-C (p < 0.05), whereas the numbers of 3609N.2, J558.47, and 7183.2 rearrangements were either lower or equivalent in Fr. B-C and Fr. C’. These data are consistent with the hypothesis that B cells expressing high IF V<sub>HI</sub> undergo more rounds of cell division at the cycling pre-B cell stage than B cells expressing low IF V<sub>HI</sub>.

Minor shifts in the VH IF fraction occur in BCR-expressing B cell subsets

To determine if BCR-mediated selection also influences the V<sub>HI</sub> IF fraction, we analyzed a high IF V<sub>HI</sub> (J606.1), an intermediate IF VH (V<sub>H</sub>10), and a low IF VH (3609N.2) at different stages of development (Fig. 4A). For J606.1 and 3609N.2, there were no statistically significant changes in the IF fraction in BCR-expressing B cell subsets. In the case of V<sub>H</sub>10, however, there were smaller but statistically significant changes that occurred beyond the cycling pre-B cell stage. To evaluate further the influence of the BCR on the V<sub>HI</sub> IF fraction, we measured the V<sub>HI</sub> IF fraction for several different VH<sub>HI</sub> in populations of splenic B cells that differed with respect to their BCR L chains (Fig. 4B). Unsorted B6 splenocytes express mostly κ L chains, whereas λ B cells and Vk8 B cells express more restricted L chain repertoires. When the V<sub>HI</sub> repertoires of these different L chain-expressing B cell populations were compared, only V<sub>H</sub>10 rearrangements exhibited a statistically significant shift in the IF fraction in the λ-expressing B cells compared with that of Vk8- or κ-expressing B cells (Fig. 4B). When the average CDR3 lengths were compared, V<sub>H</sub>10, J606.1, J558.85, and J558.72 rearrangements differed significantly between the κ-expressing splenocytes versus the more L chain-restricted B cell populations (Supplemental Fig. 3C). Taken together, these data indicate that the major shift in the IF fraction occurs between Fr. B-C and Fr. C’. Furthermore, the stability of the V<sub>HI</sub> IF fraction in BCR-expressing developmental stages and populations with different L chains suggests that the V<sub>HI</sub> IF fraction is relatively insensitive to BCR-mediated selection.

Sequence comparison of high versus low IF VHs

The preceding experiments indicate that the V<sub>HI</sub> IF fraction is determined mainly by selection at the pre-B to pre-B cell transition and may be linked to pre-B cell proliferation in cycling pre-B cells. To search for a mechanism for this selective process, we compared the amino acid sequences of high versus low IF V<sub>HI</sub>. We began by aligning all of the V<sub>HI</sub> amino acid sequences and looking for motifs that were shared among the high IF V<sub>HI</sub>. Individual V<sub>HI</sub> can be grouped on the basis of their framework sequences into larger families called clans. Clan I contains the large D-distal J558 family, as well as SM7 and V<sub>H</sub>15. Clans II and III contain all of the remaining V<sub>HI</sub> families.

We noticed that nearly all of the clan I V<sub>HI</sub> had a high IF fraction. When we used a degenerate primer to amplify several members of the J558 V<sub>HI</sub> family on B6 splenocytes, the spectratypes yielded a telltale pattern of peaks oscillating at three nucleotide intervals, with virtually no peaks in between, consistent with a very high overall IF fraction for the entire J558 family (data not shown). The only low IF J558 family member that we identified was J558.47, which had a promoter that lacked a TATA box, unlike other J558 family members (43). The only other clan I member with a low IF fraction was VH15, which had a very different amino acid sequence from all other clan I (or clan II/III) family members. When we surveyed the CDR2 sequences of different J558 family members, we observed considerable vari-
ation in the pl and hydrophobicity, but all of them appeared to have high IF fractions when analyzed in individual spectratyping studies with VH-specific primers (data not shown). Thus, nearly all of the J558 family members that were surveyed appear to have high IF VHs, despite a diversity of CDR2 sequences. As expected, based on these data, there was no statistically significant correlation between the CDR2 pl and the VH IF fraction among clan I VHs (data not shown). In contrast to clan I, the IF fractions of clan II/III VHs were far more variable. Also of note, high IF clan II/III VHs tended to have a high frequency of positively charged residues; specifically, the pls in the β-strand region within their CDR2s were high (Fig. 5 and modeling data not shown).

Discussion

The data presented in this article demonstrate that individual VHs are subjected to distinctive selection. Each VH has a characteristic fraction of IF rearrangements, and that IF fraction is reproducible in different mice. The IF fractions of 20 different VHs range from 10 to 90%. For the VHs studied in detail, the IF fraction is determined at the pre-B to pre-B cell transition, implicating the pre-BCR in VH selection.

How can the range of IF fractions in different VHs be explained? We propose that B cells expressing low IF VHs undergo fewer rounds of pre-B cell division than B cells expressing high IF VHs. The observed range of cell divisions at the pre-B stage is four to six (15), which could readily accommodate the observed range of IF fractions (Supplemental Fig. 2B, 2C). This model predicts that pre-B cells that have undergone more rounds of division will be more likely to express high IF VHs and is supported by our analysis of B cells that have undergone more rounds of division (the IF fraction of VHs increases with increasing numbers of pre-B cell divisions (the IF fraction of high IF VHs will get higher and low IF VHs will get lower)).

Individual VHs seem to have either a high or a low IF fraction. There were comparatively few VHs that had intermediate IF fractions. The simplest explanation for this seemingly bimodal distribution is that high IF VHs pair well with surrogate L chain and/or signal well through Igα and Igβ. VHs reported to function poorly in pre-BCR signaling or expression, including 7183.2, VH11, and VH12, all had low IF fractions. Furthermore, many of the low IF VHs that we surveyed had noncanonical promoter sequences. For example, the only J558 family member with a low IF fraction, J558.47, has a promoter that lacks a TATA box (43). All other members of the J558 family that we studied had high IF fractions, and the J558 family is known to exhibit the highest level of transcription of any of the mouse VH families studied (46). The J558.47 has two Pu.1 binding sites and an initiator sequence that are shared by most of the other 7183 family members (43). Similarly, Q52.8, 3609N.2, and 3609.1 have promoter sequences that either lack elements or exhibit altered spacing of elements compared with other VHs in their families with higher IF fractions (43). These promoter elements may be weak, potentially resulting in lower RNA transcript abundance and, ultimately perhaps, lower levels of H chain protein expression. If the amount of H chain protein is rate-limiting for pre-BCR assembly, higher levels of H chain proteins could result in increased pre-BCR signaling, driving more rounds of pre-B cell division.

An important unresolved question is whether the VH IF fraction is correlated with autoreactivity. In this connection, it is intriguing that most high IF clan II/III VHs have high pls in CDR2. It is possible that auto reactive or multireactive high pls clan II or III VHs are positively selected at the pre-BCR checkpoint. Positive selection of auto reactive H chains has been suggested for B-1 and fetal B cells, which tend to be enriched for the expression of certain clan II/III VHs (47–49). The enrichment of auto reactive or multireactive natural (auto)Ab specificities in the B-1 repertoire may represent an innate system of defense against commonly encountered pathogens (48, 50, 51). Perhaps these positively charged residues mediate binding to a negatively charged selecting ligand or alter the overall pre-BCR structure in such a way that binding is possible (10, 52, 53). Another possibility is that positively charged amino acids in the CDR2 of clan II/III high IF VH interact with negatively charged residues in the VpreB tail. The murine VpreB tail contains both positively and negatively charged amino acids, and it is possible that the VpreB tail forms an amphipathic helix in which one face exhibits mostly positively charged residues whereas the other face exhibits mostly negatively charged residues.

In contrast to clan II/III VHs, clan I VHs exhibit a diversity of CDR2 amino acid sequences with varying pls. Despite the variations in the CDR2 pls, nearly all clan I VHs have high IF fractions. This could mean that sequences outside of CDR2 that are conserved between different J558 family members are conferring a selective advantage upon clan I VHs. Or, perhaps J558 VHs are selected for charge motifs that can at least partially cancel one another out. For example, J558 VHs with negatively charged CDR2 sequences

![FIGURE 5](http://www.jimmunol.org/)

**FIGURE 5.** Modeling of charged motifs in clan II/III VH CDR2. A, pl analysis of the CDR2 in high versus low IF clan II/III VHs. Plotted is the mean CDR2 pl ± SD for each VH group. Each dot corresponds with a single VH sequence. Only the nonconserved 5′ portion of the CDR2 sequence is used for this analysis (see boxed regions in Fig. 5B). The high IF group (in this figure only) includes five high IF VHs and three intermediate IF VHs. The difference between the high IF versus low IF clan II/II CDR2 median pl is significant: p < 0.01 (by Mann–Whitney U test). B, Alignment of high IF versus low IF germline VH amino acid sequences. The amino acid sequences are based on the germline sequences for B6 (from Ref. 43). The CDR2 is indicated by a horizontal line, and the consensus sequence is given on top of each alignment. The motifs used for pl analysis are enclosed in boxes.
could have more positively charged CDR3 sequences and vice versa. Such a model of charge distribution among the CDRs is reminiscent of the earlier observation that Abs with anti-DNA H chains [which usually have high pIs because of the presence of arginine residues (16)] often have L chains with low pIs (54). Similarly, the selection for Abs that contain both positively charged and negatively charged CDRs typifies what Mohan and colleagues (55) have described as “hardwiring of autoreactive motifs” into the Ab repertoire. In particular, the presence of a negatively charged Asp residue at H50 can veto DNA binding by positively charged amino acids in the CDRs. By selecting for H chains with positively charged or negatively charged CDRs or, as we argue in this article, potentially both, one creates the opportunity for single Abs to bind to multiple and varied Ags, including autoantigens (56). We focused on CDR2 because it was the most variable germline-encoded sequence in the V\text{H}, but it is entirely possible that sequences outside of CDR2 are important for ligand binding or surrogate L chain interaction or both. For example, within the high IF clan I V\text{H}, there are hydrophobic residues at positions 20 and 51 and an arginine residue at position 40, which are not found in some of the low IF clan I V\text{H}.

Alternatively, rather than positive selection of high IF V\text{H} by binding to self-antigen, low IF V\text{H} could be subject to negative selection, either by V\text{H} replacement or apoptosis. With regard to V\text{H} replacement, it is interesting that most of the D-distal V\text{H} (most members of the J558 family) have high IF fractions. By virtue of their position in the H chain locus, clan I V\text{H} would be the ones most likely to replace the more D-proximal low IF clan II/III V\text{H}. As discussed earlier, it is possible that structural features of clan I V\text{H} make them more suitable as high IF V\text{H}.

Alternatively, a high IF fraction of distal V\text{H} could be explained by the timing of different V\text{H} rearrangements; D-proximal V\text{H} usually rearrange first, and by the time a distal V\text{H} has rearranged, perhaps the window of opportunity for further rearrangement is more limited, causing fewer B cells with nonproductive distal V\text{H} rearrangements to survive. With regard to apoptosis, an analysis in mice with Bcl-XL transgenes revealed an increased frequency of pro-B cells and an accumulation of nonproductive rearrangements as well as rearrangements involving D reading frame 2 (57). These data suggest that B cells have a limited time window in which H chain rearrangements are permitted prior to death (they have a high crash factor). In a subsequent study, the same group showed that self-reactive Bcl-XL transgenic B cells were able to escape death but were regulated by other pathways that enforced self-tolerance, namely anergy and receptor editing (58).

Curiously, we were able to recover repeatedly two sequences from independent PCR amplifications of V\text{H} J558.85–I\text{H}2 rearrangements in pro-B cells (Supplemental Fig. 1C). This finding is perplexing as pro-B cells are not known to proliferate. These sequences were not observed in water or fibroblast DNA, but we cannot unequivocally rule out the possibility of PCR contamination. We did not observe clonal expansions of any other V\text{H} rearrangement in pro-B cells. The mechanism by which these sequences arise is unclear, but we considered the possibility of homology-mediated joining, given the findings of Chukwuocha et al. (59), who observed that a V\text{H} in the S107 family exhibited an elevated frequency of IF rearrangements in cytoplasmic Ig-negative pre-B cells from newborn mice. However, in the case of the rearrangements observed in this study, they were observed in adult mice, and there is no obvious sequence homology between J558.85 and D\text{H}DSP2.9 or an alternation in the RSS of J558.85 that readily explains these findings.

Perhaps the most intriguing property of the V\text{H} IF fraction is that it is nearly entirely established by the cycling pre-B cell stage: for most V\text{H}, the IF fraction did not shift nearly as much beyond the cycling pre-B cell stage as it did before it. Furthermore, the IF fraction of most V\text{H} does not differ significantly in \(\lambda\) versus \(\kappa\) expressing B cells. The lack of large differences in IF fraction in more mature B cell subsets suggests that the repertoire of H chains is either so dominant and/or so large that many different rearrangements of a given H chain can preserve the IF fraction, even when there is considerable narrowing of the repertoire. Narrowing of the repertoire beyond the naive B cell stage is strongly supported by the reduced level of L chain diversity in mature recirculating compared with naive BM B cells in a J558 H chain transgenic mouse model (60). Despite narrowing with regard to particular H plus L chain pairs, the Ab repertoire can nevertheless be extremely diverse. The elegant study of Owen et al. (61) demonstrated massive clonotypic diversity in the IgM response to phosphorylcholine, despite this response being restricted to only a few V\text{H}. Thus, the H chain repertoire is established early during B cell development yielding the IF fraction or selection set-point for each V\text{H}. From a nearly limitless diversity of H chains, Ab specificities are subsequently narrowed to a more “useful” array of protective specificities in more mature B cell stages based on H plus L chain pairing and further fine-tuning of Ag specificity via somatic mutation.

Overall, our data suggest that the H chain IF fraction can be used as a measure of H chain selection at the pro-B to pre-B cell transition. An altered IF fraction could point to an early B cell tolerance checkpoint defect or to altered early B cell development. B cell tolerance checkpoint defects have been observed in animal models and human subjects with lupus and type 1 diabetes as early as the L chain rearrangement stage in pre-B cells (62, 63). The H chain IF fraction potentially surveys an even earlier stage of B cell development and selection. It is intriguing that certain V\text{H} such as V\text{H}4-34, V\text{H}13-23, V\text{H}1-69, V\text{H}3-30, and others are overrepresented in humans with autoimmune disease or in polyreactive Abs (64–69). It will be interesting to determine if the V\text{H} IF fraction for these and other V\text{H} is different in individuals with autoimmunity, potentially providing insights into the stage of B cell development where immune tolerance is defective in individual patients with autoimmunity. A measure of early B cell tolerance or development, such as the V\text{H} IF fraction, may be helpful in targeting therapy for autoimmunity to the relevant B cell subset(s). In addition, an abnormal IF fraction of a single germline V\text{H} could reveal a more specific change in the Ab repertoire, such as clonal expansion in response to a particular (auto)antigen (70).

Acknowledgments

We thank the University of Pennsylvania Flow Cytometry core facility and the molecular genotyping facilities at the Hospital of the University of Pennsylvania and the Department of Genetics core laboratory at the University of Pennsylvania for skilled technical assistance. We thank Siyuan Hu for expert help with illustrations. We also gratefully acknowledge Michael Denny, Avinash Bhandoola, Michael Cancro, Robert Wilson, Olga Kalinina, Sara Smith, Lee Herzenberg, Robert F. Thomas, Jr., and Denise Gay for helpful discussions.

Disclosures

The authors have no financial conflicts of interest.

References