Regulation of Repertoire Development through Genetic Control of $D_H$ Reading Frame Preference


*J Immunol* 2008; 181:8416-8424; doi: 10.4049/jimmunol.181.12.8416
http://www.jimmunol.org/content/181/12/8416

---

**References**
This article cites 31 articles, 16 of which you can access for free at:
http://www.jimmunol.org/content/181/12/8416.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
Regulation of Repertoire Development through Genetic Control of $D_H$ Reading Frame Preference


In jawed vertebrates most expressed Ig H chains use only one of six possible $D_H$ reading frames. Reading frame (RF)1, the preferred reading frame, tends to encode tyrosine and glycine, whereas the other five RFs tend to be enriched for either hydrophobic or charged amino acids. Mechanisms proposed to favor use of RF1 include a preference for deletion over inversion that discourages use of inverted RF1, RF2, and RF3; sequence homology between the 5′ terminus of the $J_H$ and the 3′ terminus of the $D_H$ that promotes rearrangement into RF1; an ATG start site upstream of RF2 that permits production of a truncated $D_\mu$ protein; stop codons in RF3; and, following surface expression of IgM, somatic, presumably Ag receptor-based selection favoring B cells expressing Igs with tyrosine- and glycine-enriched CDR-H3s. By creating an IgH allele limited to the use of a single, frameshifted $DFL16.1$ $D_H$ gene segment, we tested the relative contribution of these mechanisms in determining reading frame preference. $D_\mu$-mediated suppression via an allelic exclusion-like mechanism dominated over somatic selection in determining the composition of the CDR-H3 repertoire. Evidence of somatic selection for RF1-encoded tyrosine in CDR-H3 was observed, but only among the minority of recirculating, mature B cells that use $D_H$ in RF1. These observations underscore the extent to which the sequence of the $D_H$ acts to delimit the diversity of the Ab repertoire. The Journal of Immunology, 2008, 181: 8416–8424.

Among the six CDRs that make up the Ag binding site in Igs, it is the CDR of the H chain that most often plays a defining role in Ag recognition and binding (1–3). This key role reflects the location of CDR-H3 at the center of the Ag binding site, as classically defined, as well as the two rounds of N addition permitted by the inclusion of a diversity (D) gene segment (4, 5).

In theory, the inclusion of a D gene segment coupled with random insertion of N nucleotides should produce a CDR-H3 repertoire of random diversity. This permissive role of the D has been referred to as D-diversity (6). In practice, tyrosine and glycine are heavily overrepresented in CDR-H3, comprising 30–40% of the global amino acid content of this hypervariable interval. Many of these tyrosines and glycines are provided to CDR-H3 by use of $D_H$ reading frame 1 (RF1), which is normally favored (7–15). This preference for RF1 has the effect of disenfranchising almost two-thirds of all DJ rearrangements, which at first glance appears extremely wasteful. This has led to the suggestion that the expression of $D_H$ in RF2 might be incompatible with effective Ag recognition and downstream B cell signaling, a concept referred to as D-disaster (6).

Mechanistically, the normal preference for RF1 has been linked to RF-specific properties and sequence motifs that are shared among 12 of the 13 $D_H$ gene segments in BALB/c mice. These include a predilection for rearrangement by deletion; the frequent occurrence of stop codons in RF3, which act to reduce the likelihood of creating an open reading frame (ORF) among VDJ rearrangements that use RF3; a bias toward rearrangement at sites of sequence microhomology between the 5′ end of the $J_H$ and the 3′ end of the $D_H$, which favor rearrangement into RF1; and an ATG start site upstream of RF2 that permits production of a truncated $D_\mu$ protein (9, 16–19). The $DQ52$ gene segment, which contributes to <5% of the repertoire, shares none of these features and use of RF1 is less pronounced.

Transgenic studies have shown that the bias against RF2 can be released when pre-B cells are no longer able to produce membrane-bound $D_\mu$ protein, suggesting that this $\mu$ chain variant can engage the mechanisms of allelic exclusion to inhibit subsequent V→DJ rearrangement (20). However, the extent to which the bias for tyrosine and glycine in CDR-H3 reflects genetic control of $D_H$ reading frame rearrangement preferences remains unclear, as does the extent to which somatic selection during B cell development might adjust the repertoire should use of RF2-encoded amino acids yield a D-disaster (6).

Received for publication June 23, 2008. Accepted for publication October 8, 2008.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This work was supported in part by National Institutes of Health Grants AI07051, AI42732, AI48115, HD043327, and T020130; by Deutsche Forschungsgemeinschaft SFB/TR22-TPA17; and by Alexander von Humboldt-Stiftung FFL1071857.

2 M.Z. and R.L.S. contributed equally to this work.

3 Address correspondence and reprint requests to Dr. Harry W. Schroeder, Jr., Departments of Medicine, Microbiology, Genetics, Pediatric Dentistry, and Pediatrics, University of Alabama at Birmingham, Shelby Building 176, 1530 3rd Avenue South, Birmingham, AL 35294. E-mail address: hwsj@uab.edu

4 Abbreviations used in this paper: RF, $D_\mu$ reading frame; $\Delta D$-DFL, depleted $D_H$ locus with a single $DFL16.1$ gene segment; $\Delta D$-iD, depleted $D_H$ locus with a single, mutated $DFL16.1$ gene segment containing inverted DSP2.2 sequence; $\Delta D$-DuFS, depleted $D_H$ locus with a single, frameshifted $DFL16.1$ gene segment; CDR-H3, Ig heavy-chain CDR3; ES, embryonic stem; ORF, open reading frame; RSS, recombination signal sequence.

Copyright © 2008 by The American Association of Immunologists, Inc. 0022-1767/08/$2.00
To address the role of D<sub>H</sub> sequence in regulating RF usage and CDR-H3 amino acid content, we evaluated the composition of the CDR-H3 repertoire among developing B cells in the bone marrow of mice limited to use of a single, mutated, frameshifted D<sub>H</sub> DFL16.1. The first of two mutations shifts the D<sub>H</sub> ORF from RF2 to RF1. The second shifts the region of D<sub>H</sub>-J<sub>H</sub> microhomology from RF2 to RF1 and at the same time shifts one of two TAG codons from RF3 to RF1. We show that these two frameshift mutations completely transfer the preference for RF1 to RF2 among progenitor B cells. Although we do find evidence of selection for D<sub>H</sub> RF1 sequence content in mature B cells, this effect appears limited to that minority of the mature B cell population that uses RF1. For most cells, we show that D<sub>H</sub>-mediated suppression via an allelic exclusion-like mechanism dominates over D<sub>H</sub> sequence composition in shaping the diversity of the CDR-H3 repertoire in mice. Our results support the view that the mature B cell population contains over D<sub>H</sub> sequence composition in shaping the diversity of the CDR-H3 repertoire.

### Materials and Methods

**Generation of targeted embryonic stem (ES) cells and the ΔΔ-D<sub>H</sub>FS mouse**

The methods used to create the IgH ΔΔ-D<sub>H</sub>FS allele were identical with those previously used to create the ΔΔ-D<sub>FL</sub>FL and ΔΔ-idos alleles (14, 15) with the following exception. An 800-bp BglII fragment containing DFL16.1, a gift from Dr. Y. Kurosawa (Fujita Health University, Japan) (8), was modified by PCR-based site-directed mutagenesis to include a thymidine nucleotide at the 5' end (8), was modified by PCR-based site-directed mutagenesis to include a thymidine nucleotide at the 5' end from the 5' end, see Fig. 1) (21). The ESDQ52-KO cell line, which was derived from wild-type BALB/c-1 ES cells of the Ig<sub>H</sub> haplotype (22), was again targeted. This cell line contains a loxP site in lieu of a 240-bp XhoI-Sacl fragment containing the DQ52 gene and a putative 5' cis-regulatory element. By Southern blot analysis and DNA sequencing, one ES clone was identified to contain both the D<sub>H</sub>FS- and the DQ52-deleted mutations. Following injection into C57BL/6J blastocysts, a chimeric D<sub>H</sub>FS- containing male was bred to a BALB/c H-2<sup>d</sup> transgenic female. All experiments with live mice were approved by and performed in compliance with Institutional Animal Care and Use Committee regulations.

**Flow cytometric analysis and cell sorting**

Flow cytometric analysis and cell sorting were performed as previously described on cells from bone marrow (12, 14, 15). A MoFlo instrument (Dako) was used for cell sorting. Cells were independently sorted from the bone marrow of two homozygous ΔΔ-D<sub>H</sub>FS and two wild-type siblings (8 wk old). Developing B lineage cells in the bone marrow were identified on the basis of surface CD19, CD43, IgM, BP-1, and IgD (see Fig. 2) (23).

**RNA, RT-PCR, cloning, sequencing, and sequence analysis**

Total RNA isolation, V<sub>3</sub>7183-specific VDJ<sub>μ</sub> RT-PCR amplification, cloning, sequencing, and sequence analysis was performed as previously described (12, 14, 15). The sequences reported in this paper have been placed in the GenBank database (accession nos. EF627050–EF627448). A listing of the distribution by developmental stage of the 1355 wild-type, 235 ΔΔ-D<sub>FL</sub>FS, 243 ΔΔ-D<sub>H</sub>FL, and 342 ΔΔ-ids sequences used for analysis in the work is provided in supplemental Table S1. The CDR-H3 sequences obtained from the ΔΔ-D<sub>H</sub>FS mice and their wild-type littermate controls are presented in disassembled form in an accompanying Microsoft Excel file labeled “Supplemental Compilation of CDR-H3 Sequences”.

**Statistical analysis**

Statistical analysis was performed with JMP version 7.0 (SAS Institute) as previously described (12, 14, 15). Means are accompanied by the SEM.

### Results

**Frame shifting the D<sub>μ</sub> ATG start site and the D<sub>H</sub>-J<sub>H</sub> microhomology region transfers the reading frame bias of DFL16.1 from RF1 to RF2**

To test the role of D<sub>H</sub> sequence in the regulation of reading frame usage, we introduced two frameshift mutations into the sequence of DFL16.1. By inserting a thymidine at the 5' terminus of DFL16.1, the first mutation shifted the D<sub>μ</sub> ORF from RF2 to RF1 (Fig. 1, frameshift no. 1). By inserting a thymidine six bases upstream of the 3' terminus of DFL16.1, the second mutation shifted the region of D<sub>H</sub>-J<sub>H</sub> microhomology from RF1 to RF2 (Fig. 1, frameshift no. 2). This second mutation also shifted the distal RF3-encoded TAG termination codon from RF3 to RF1. We termed the resulting D<sub>H</sub> mutant D<sub>H</sub>FS, for D<sub>H</sub> frameshift (21).

To introduce the D<sub>H</sub>FS allele into the BALB/c genome, we modified an existing DFL16.1 loxP-neo-loxP construct (14, 15) to encode D<sub>H</sub>FS in place of wild-type DFL16.1. We introduced D<sub>H</sub>FS into the DFL16.1 locus of a BALB/c ES cell that had previously undergone cre-loxP-mediated removal of J<sub>H</sub>-proximal DQ52 (22). To delete (ΔD) the intervening 11 D<sub>H</sub> and the neo-loxP cassette, we bred a heterozygous BALB/c D<sub>H</sub>FS male mouse to a BALB/c cre-expressing female. The offspring whose D<sub>μ</sub> locus had undergone deletion between the DFL16.1-proximal and the J<sub>H</sub>-proximal loxP sites retained only D<sub>H</sub>FS in this new ΔΔ-D<sub>H</sub>FS IgH<sup>α</sup> allele. We bred this new mutant D<sub>H</sub> into homozygosity.

![Figure 1](http://www.jimmunol.org/) - Two frameshift mutations were used to shift D<sub>H</sub> RF bias from RF1 to RF2. In the targeting construct, a thymidine was inserted at the 5' terminus of DFL16.1, the most V<sub>3</sub>-proximal D<sub>H</sub> (33). This insertion placed RF1 in frame with the upstream D<sub>μ</sub> ATG start site (20). A second thymidine was inserted six bases upstream of the 3' terminus of DFL16.1. This insertion placed RF2 in frame with terminal codons for serine and tyrosine that share sequence with the 5' terminus of three of the four J<sub>H</sub> gene segments (7, 9).
Flow cytometric gates for the collection of bone marrow B lineage cells from homozygous ΔD-DμFS and wild-type (WT) littermates. Cells within the lymphocyte gate were first distinguished on the basis of the expression of CD19 and CD43. Early B cell progenitors (CD19⁺CD43⁻) were divided into fractions B and C on the basis of the expression of BP-1. Late pre-B, immature B, and mature B cells (CD19⁺CD43⁺) were divided into fractions D, E, and F on the basis of the surface expression of IgM and IgD.

FIGURE 2.
FIGURE 3. The two ΔH frameshift mutations flip the usage ratio of RF1 vs RF2. The percentage usage of RF1 and RF2 among unique, in-frame Igμ transcripts with recognizable DFL16.1 sequence from Hardy fractions B (CD19⁺CD43⁺IgM⁻BP-1⁻), C (CD19⁺CD43⁺IgM⁺BP-1⁻), D (CD19⁺CD43⁺IgM⁺IgD⁺), E (CD19⁺CD43⁺IgM⁺IgD⁻), and F (CD19⁺CD43⁻IgM⁻IgD⁻) from wild-type, ΔD-DFL, and ΔD-ΔμFS mice is displayed. When compared with ΔD-DFL and wild-type controls, the use of hydrophobic and neutral DH reading frames has been flipped in B lineage cells from ΔD-ΔμFS mice.

To test whether these ΔH frameshift mutations would alter RF preference early in B cell development, we used the scheme of Hardy and Hayakawa (23) to identify B cell progenitors from the bone marrow of homozygous BALB/c ΔD-ΔμFS mice. We RT-PCR cloned Vμ7183-D-J-Cμ transcripts from fraction B (CD19⁺CD43⁺BP-1⁻IgM⁻ pro-B cells) (Fig. 2), a stage before that associated with cell surface expression of Igμ. In these and other experiments, we compared the results obtained from sequencing Vμ7183-D-J-Cμ transcripts from ΔD-ΔμFS mice to those of similar transcripts from both wild-type and homozygous ΔD-DFL BALB/c mice. The latter are restricted to use of only DFL16.1 in their ΔH locus (15). When compared with ΔD-DFL and wild-type mice, the ratio of DFL16.1 RF1 to RF2 usage was reversed (Fig. 3).

The frequency of RF3, which retained its central termination codon, did not change. Thus, we found no in vivo evidence that the elimination of a terminal stop codon in RF3, or the introduction of a terminal stop codon to RF1, would affect the prevalence of RF3-containing CDR-H3s among Igμ sequences with ORFs.

Given that the distal TAG from RF3 now served as the penultimate codon for RF1 (Fig. 1), we calculated that an exonucleolytic loss of at least five nucleotides at the 3’ terminus of the DμFS gene segment would be required to permit formation of an ORF in RF1-containing transcripts. Unsurprisingly, on average fraction B DμFS RF1-containing ORF transcripts lost eight 3’ nucleotides, whereas the average 3’ loss in RF2-containing ORFs was only four (p < 0.001).

Enhanced use of RF2 yields a hydrophobic CDR-H3 repertoire enriched for valine

The new bias for use of RF2 among the B cell progenitors in fraction B from the ΔD-ΔμFS mice translated into a globally altered polyclonal CDR-H3 repertoire enriched for RF2-encoded hydrophobic amino acids (Fig. 4). In particular, translation resulted in enrichment for RF2-encoded valine with a corresponding depletion of RF1-encoded tyrosine and glycine.

Terminal ΔH-JH microhomology appears to play a limited role in controlling ΔH RF preference in ΔD-ΔμFS mice

The role of ΔH-JH microhomology would be predicted to be most apparent among rearrangements that lack N nucleotides at the D→J junction, especially among those with a demonstrable overlap in DJ sequence. The number of sequences from fraction B that fit these descriptions was insufficient to allow meaningful analysis. However, to pursue our studies of the role of the ΔH in repertoire development, we also obtained CDR-H3 sequences from in-frame Vμ7183-D-J-Cμ transcripts from fraction C (CD19⁺CD43⁺BP-1⁻IgM⁻ early-pre-B cells), fraction D (CD19⁺CD43⁻IgM⁺IgD⁻ late-pre-B cells), fraction E (CD19⁺CD43⁻IgM⁺IgD⁺ immature B cells), and fraction F (CD19⁺CD43⁻IgM⁺IgD⁻ mature, recirculating B cells) (Fig. 2). We grouped all of the sequences that we obtained and then examined the pattern of RF1 among those lacking nucleotides between ΔH and JH with a special focus on those sequences with evidence of one or more nucleotides of DJI overlap. We compared the results obtained in the ΔD-ΔμFS mice to the ΔD-DFL control, which, because it also contained only one ΔH, also allowed identification of N nucleotides with certainty.

As predicted, among the ΔD-DFL sequences the frequency of rearrangement into RF1 was greater among those DJ joins that lacked N nucleotides (49 of 60 (82%)) than among the DJ joins that contained them (111 of 164 (68%), p < 0.05). Unexpectedly, the same did not hold true for ΔD-ΔμFS. RF2 usage was nearly identical between those sequences that lacked N nucleotides (31 of 44 (70%)) and those that contained them (112 of 165 (68%); p = 0.74). This same trend was evident at all stages of B cell development examined, although the numbers were too small to reach statistical significance for each individual stage.

Closer inspection revealed an even greater shift in RF usage among the ΔD-ΔμFS sequences with direct DJI overlap, with only half (8 of 16) using RF2 vs 93% of the ΔD-DFL sequences using RF1 (28 of 30, p < 0.002). This reflected an altered pattern of the site of rearrangement at the 3’ terminus of the D. Whereas the average loss of 3’ nucleotides in DFL16.1 sequences among the wild-type and ΔD-DFL sequences that contained DJI overlap was statistically indistinguishable (0.8 ± 1.3 and 1.9 ± 0.7, respectively; p = 0.45), the average DμFS sequence containing DJI overlap was more than twice that (4.8 ± 1.0, p < 0.03 for both controls). Thus, a frameshift in the midst of the coding sequence of the ΔH altered the behavior of the rearrangement process.
No major evidence of selection for use of RF1 or for RF1-encoded amino acids

To test whether B cell maturation, which is dependent in part on successful testing of the Ag-binding properties of the BCR, would be associated with selection for that minority of developing B cells that had rearranged their \( D_H \) in RF1, we evaluated RF usage among in-frame VH7183-D-J-C\( _{\mu} \) transcripts obtained from fraction C through fraction F bone marrow B lineage cells (Fig. 2). Not only did we fail to find enrichment for RF1 with development, we actually observed enrichment for RF2 (Fig. 3). The increase in the use of RF2 with B cell maturity matched that observed for RF1 among both \( \Delta D-DFL \) and wild-type controls.

To test whether B cell maturation would be associated with selection for tyrosine and glycine, which are normally preferred, we examined global amino acid content in the CDR-H3 loop sequences from fraction C to F. The new bias for RF2-encoded hydrophobic amino acids persisted with development up to and including fraction F (Fig. 4A), persistently skewing the distribution of average CDR-H3 hydrophobicity in this recirculating, mature B cell population toward the hydrophobic end of the spectrum (Fig. 4B). In particular, we found no evidence of selection for RF2-containing CDR-H3s that had gained tyrosine or glycine by means of N region addition coupled with exonucleolytic loss of terminal \( D_H \) sequence.

In wild-type mice, sequences with an average hydrophobicity greater than +0.60 or less than −0.70 are progressively culled from the repertoire during B cell development (12, 13). Among those sequences obtained from \( \Delta D-\mu FS \) fraction F B cells, several exhibited CDR-H3s with an average hydrophobicity greater than +0.60 (Fig. 4B). Coupled with a failure to enrich for sequences within the normally preferred hydrophobic range of more than −0.40 and ≤0.00, the presence of these hydrophobic sequences led to an increase in average CDR-H3 loop...
Each DH reading frame appears to generate its own characteristic global amino acid repertoire

To further test for evidence of selection for or against specific patterns of amino acid usage of the global CDR-H3 repertoire, we compared CDR-H3 composition by RF. The numbers of sequences were insufficient to allow meaningful analysis by developmental stage. However, given that the overall pattern of amino acid content appeared relatively stable (Fig. 4A) (15), we reasoned that comparing RF1 vs RF2 sequences regardless of developmental stage would be appropriate. We found that the ΔD-DμFS RF2 repertoire generally matched that generated by DFL16.1 RF2 in both wild-type and ΔD-DFL mice (data not shown). Similarly, RF1 in the ΔD-DμFS mice generated a neutral CDR-H3 loop repertoire with a pattern of amino acid usage similar to that obtained from the RF1-generated repertoire from controls.

Selection for length and D<sub>μ</sub> content among DμFS CDR-H3 sequences that use RF1

In both mice and humans, a selection for a specific range and average of CDR-H3 lengths is consistently observed during B cell development (12, 13, 24). As noted above, making TAG (Fig. 1) the penultimate 3′ codon of DμFS RF1 would require that this mutant D<sub>μ</sub> undergo a minimum loss of five nucleotides to produce a functional V domain. This was borne out in practice, with productive in-frame V<sub>μ</sub>7183-D-J-C<sub>μ</sub> transcripts from ΔD-DμFS fraction B cells losing an average of six 3′ terminal nucleotides vs an average loss of only four nucleotides among transcripts from ΔD-DFL mice (data not shown). Similarly, RF1 in the DFL16.1 RF2 repertoire generally matched that generated by DFL16.1 RF2 in both wild-type and ΔD-DFL mice (data not shown). Similarly, RF1 in the ΔD-DμFS mice generated a neutral CDR-H3 loop repertoire with a pattern of amino acid usage similar to that obtained from the RF1-generated repertoire from controls.

Note that a complete listing of the nucleotide and predicted amino acid sequences of the ΔD-DμFS CDR-H3s reported in this paper can be found in the accompanying supplemental Microsoft Excel file.

No major changes in V<sub>μ</sub> and J<sub>μ</sub> usage

To test whether the shift in CDR-H3 hydrophobicity might be “tempered” by a compensatory change in V<sub>μ</sub> or J<sub>μ</sub> usage, we compared V<sub>μ</sub>7183 and J<sub>μ</sub> usage in the ΔD-DμFS fraction B lineage cells to controls. The pattern of use of V<sub>μ</sub>7183 gene segments proved statistically indistinguishable from the controls (12, 14, 15) (Fig. 5). When compared with wild type, the use of J<sub>μ</sub>4 was diminished and use of J<sub>μ</sub>1 was enhanced among ΔD-DμFS-containing transcripts (Fig. 5). This pattern matched that observed in the CDR-H3 repertoires of the two previous single D<sub>μ</sub> mice (14, 15) and likely reflects the elimination of the rest of the D<sub>μ</sub> locus, including DQ52 (22).

Each D<sub>μ</sub> reading frame appears to generate its own characteristic global amino acid repertoire

To further test for evidence of selection for or against specific patterns of amino acid usage of the global CDR-H3 repertoire, we compared CDR-H3 composition by RF. The numbers of sequences were insufficient to allow meaningful analysis by developmental stage. However, given that the overall pattern of amino acid content appeared relatively stable (Fig. 4A) (15), we reasoned that comparing RF1 vs RF2 sequences regardless of developmental stage would be appropriate. We found that the ΔD-DμFS RF2 repertoire generally matched that generated by DFL16.1 RF2 in both wild-type and ΔD-DFL mice (data not shown). Similarly, RF1 in the ΔD-DμFS mice generated a neutral CDR-H3 loop repertoire with a pattern of amino acid usage similar to that obtained from the RF1-generated repertoire from controls.

Selection for length and D<sub>μ</sub> content among DμFS CDR-H3 sequences that use RF1

In both mice and humans, a selection for a specific range and average of CDR-H3 lengths is consistently observed during B cell development (12, 13, 24). As noted above, making TAG (Fig. 1) the penultimate 3′ codon of DμFS RF1 would require that this mutant D<sub>μ</sub> undergo a minimum loss of five nucleotides to produce a functional V domain. This was borne out in practice, with productive in-frame V<sub>μ</sub>7183-D-J-C<sub>μ</sub> transcripts from ΔD-DμFS fraction B cells losing an average of eight 3′ terminal nucleotides vs an average loss of only four 3′ terminal nucleotides among DFL16.1-containing transcripts from the controls (data not shown). Because the extent of 5′ loss among the ΔD-DμFS fraction B cells was statistically indistinguishable from controls (Fig. 6B), the increased loss of 3′ nucleotides led to ΔD-DμFS RF1 contributing, on average, ~four fewer germline-encoded nucleotides to CDR-H3 than did controls (p < 0.05) (Fig. 6C). As a result, the average length of CDR-H3 in fraction B was almost one codon shorter than controls (Fig. 6D).

With development, however, the contribution of germline D<sub>μ</sub> sequence increased. This effect was most notable among B cells from fraction F, where the average loss of 3′ nucleotides had dropped from eight to seven, and, more importantly, where the average loss of 5′ nucleotides dropped precipitously from six to two (p < 0.05). The effect was to create a mature ΔD-DμFS RF1 repertoire that maintained the same relative contribution of germline D<sub>μ</sub> sequence as wild-type DFL16.1 RF1 in CDR-H3 intervals of the same average length. Although the numbers were insufficient to perform a statistical analysis of global amino acid content, in general the effect of the change in nucleotide loss would be to compensate for the loss of 3′ tyrosine with a 5′ tyrosine encoded by the 5′ terminus of the D<sub>μ</sub> (Fig. 1). Thus, while the absolute sequence of the RF1-encoded CDR3-centric Ag binding site repertoire was not being recreated, the global amino acid composition of the repertoire in mature B cells was more likely to match that of wild-type controls.
Discussion

Although D<sub>h</sub> gene segments were named for their role in increasing the potential for diversity at the time of Ig rearrangement, that is, D-diversity, they conserve in their sequence a number of molecular properties that have been proposed to regulate both choice of RF and the amino acid content of CDR-H3, which forms the center of the classic Ag binding site (10, 20). In the present work, we have taken a genetic approach to directly test the role of several molecular properties that have been proposed to regulate both choice of RF and the amino acid content of CDR-H3, including the spacer and the adjacent coding sequence, can influence patterns of rearrangement (26–28). In our case, the mutations we created maintained the adjacent coding sequence, can influence patterns of rearrangement (26–28). With maturation, there is increased preservation of both 3′ and 5′ terminal nucleotides among the transcripts that contain the D<sub>h</sub>FS gene segment. This achieves near identity in the contribution of germline-encoded D<sub>h</sub> nucleotides to the mature B cell CDR-H3 repertoire (25), and that the sequences between the heptamer and nonamer can be conserved within groups of evolutionarily related sequences (25), and that the sequences flanking the RSS heptamer and nonamer, including the spacer and the adjacent coding sequence, can influence patterns of rearrangement (26–28). In our case, the mutations we created maintained the germline sequence and context of the RSS and adjacent nucleotides.

A number of previous studies have provided strong support for the role of terminal D/J microhomology in biasing RF preference (9, 18, 29). We designed the 3′ insertion to shift the region of terminal D<sub>h</sub> RF1-J<sub>h</sub> microhomology to RF2. Because the sequence of the J<sub>h</sub> locus remained unchanged, the putative ability of the region of D<sub>h</sub>-J<sub>h</sub> microhomology to contribute to the coding sequence of CDR-H3 was maintained. However, shifting the six terminal 3′ nucleotides of DFL16.1 from RF1 to RF2 appeared to have a minimal effect on RF2 usage in the developing repertoire. Given that the rest of the adjacent sequence of DFL16.1 was maintained in germline content and that the actual site of rearrangement in sequences with demonstrable D/J overlap changed as a result of the frameshift, it remains possible that the coding sequence adjacent to the region of terminal homology also exerts an effect on processing of the terminal end of the D<sub>h</sub> during D→J rearrangement and thus on RF bias. This effect would necessarily be separate from the direct role of terminal D/J microhomology.

RF3 in DFL16.1 normally contains two TAG termination codons (Fig. 1). The insertion of thymine 7 bp for the 3′ end of the coding sequence moved the distal TAG termination codon from RF3 to
RF1. We hypothesized that the inclusion of a termination codon in RF1 would lead to a further decrease in its use in the functional repertoire due to the decreased likelihood of creating ORFs. We further assumed that the loss of a termination codon in RF3 would lead to an increase in its representation in the functional repertoire. However, the prevalence of the modified RF3, which retained its first, central termination codon, remained unchanged. Furthermore, the prevalence of ΔD-DµFS RF1, which now encodes a TAG as its penultimate codon, matched that observed in RF2 in both the wild-type and ΔD-DFL controls. This suggests that the relative position of the termination codon in a D_H RF may be critical to its ability to limit use of that RF.

The 5′ insertion shifted the upstream Dµ ATG start site from RF2 to RF1. Previous studies have shown that disruption of the Cµ membrane terminus leads to random representation of D_H RFs (20). These experiments linked expression of the membrane form of Dµ protein with inhibition of V→DJ joining and provided initial evidence that the bias in D_H usage was the product of repertoire selection on an evolutionary scale. Subsequent studies in mice unable to express the surrogate L chain molecule λ5 (30) provided further support for the hypothesis that selection against RF2 occurred primarily in pre-B cells by engaging the process of allelic exclusion.

Our studies now show that moving the RF1 coding sequence in-frame with the Dµ ATG start site increased the representation of DFL16.1 RF2 among the CDR-H3 repertoire in progenitor B cells to that normally observed for RF1, and vice versa. Thus, of the three mechanisms postulated to regulate RF usage that we tested, the negative effect of a Dµ ORF appeared to dominate, further confirming the key role of Dµ protein expression in regulating D_H RF preference in mice.

Preservation of the pattern of amino acid representation by RF is a major conserved feature of diversity gene segments (10). In previous studies we have shown that we could completely alter the general amino acid composition of the CDR-H3 repertoire expressed by progenitor B cells to that normally observed for RF1, and vice versa. Thus, the three mechanisms postulated to regulate RF usage that we tested, the negative effect of a Dµ ORF appeared to dominate, further confirming the key role of Dµ protein expression in regulating D_H RF preference in mice.

For mice bearing the ΔD-dm mutation (14), recreation of a more wild-type tyrosine and glycine-enriched repertoire would have required selection for rare sequences that either contained a D_H gene segment that had undergone inversion or that had undergone extensive exonucleolytic loss coupled with just the right combination of N nucleotides to recreate a tyrosine and glycine-enriched sequence. In contrast, for mice bearing the ΔD-DµFS mutation, one-fifth of the transcripts in the progenitor B cell population were already using the normally preferred RF1. This should have lowered the threshold for detection of somatic selection during development for tyrosine and glycine-enriched CDR3-centric Ag binding sites.

Examination of the repertoire expressed by pre-B cells, immature B cells, and mature B cells revealed that the switch from RF1 to RF2 dictated to progenitor cells by inverting the coding sequence of DFL, which is subject to natural selection, plays a key role in regulating, or delimiting, the composition of CDR-H3. The extent of that role, however, remained uncertain.

For mice bearing the ΔD-dm mutation (14), recreation of a more wild-type tyrosine and glycine-enriched repertoire would have required selection for rare sequences that either contained a D_H gene segment that had undergone inversion or that had undergone extensive exonucleolytic loss coupled with just the right combination of N nucleotides to recreate a tyrosine and glycine-enriched sequence. In contrast, for mice bearing the ΔD-DµFS mutation, one-fifth of the transcripts in the progenitor B cell population were already using the normally preferred RF1. This should have lowered the threshold for detection of somatic selection during development for tyrosine and glycine-enriched CDR3-centric Ag binding sites.

Examination of the repertoire expressed by pre-B cells, immature B cells, and mature B cells revealed that the switch from RF1 to RF2 dictated to progenitor cells by the altered molecular properties of the mutated D_H persisted as the developing B cells passed through subsequent developmental checkpoints even though this created a global shift in CDR-H3 amino acid content. We found no evidence for selection of RF2 rearrangements that had undergone enhanced N addition or increased exonucleolytic loss of D_H sequence, either of which could have reduced use of valine or increased the contribution of tyrosine and glycine in sequences that used RF2. Finally, there was no demonstrable change in V_H or J_H usage other than that imposed by the reduction in available D_H elements. Taken together, these findings underscore the critical role of germline D_H sequence in preselecting the diversity of the preimmune CDR-H3 repertoire.

Although the results of this study support a dominant role for germline D_H sequence in delimiting the general global amino acid content of the CDR-H3 repertoire, closer inspection revealed a subtle, but significant, effect of selection for a specific range of germline-encoded CDR-H3s into the mature, recirculating B cell pool. Despite the forced loss of two terminal codons, the DFL16.1 RF1-containing CDR-H3 repertoire in mature, recirculating B cells was observed to match the extent of germline D_H nucleotide contribution and CDR-H3 length found in both the wild-type and ΔD-DFL controls. The finding that there is somatic selection among the 20% of sequences that use DµFS in RF1 but not among the 60% of sequences that use DµFS in RF2 suggests separate regulation by RF of CDR-H3 content. This would suggest that exposure to common Ags in the periphery selects for a specific range of DFL16.1 RF1-encoding CDR-H3s.

In previous studies we have shown that mice forced to express an Ab repertoire depleted of CDR-H3 tyrosine and glycine, but enriched for arginine and other charged amino acids, demonstrate reductions in B cell and Ab production as well as increased susceptibility to infection with a common pathogenic virus (14, 31). In a related series of studies, we have found that enrichment for RF2-encoded valine-enriched CDR-H3s can also degrade B cell development and Ag-specific Ab production (32), even when a significant minority of B cells continues to produce a tyrosine and glycine-enriched CDR-H3 repertoire. This would suggest that failure to properly maintain the correct balance of amino acids within the CDR-H3 repertoire as a result of chance, age, or an inherited tendency could alter immune function in otherwise healthy patients, potentially influencing susceptibility to disease. These data also support the view that the evolution of D gene sequence may reflect a balance between the need to avoid the D-disaster (6) that can occur with use of disfavored D_H sequence and the need to maintain the D-diversity necessary to create an effective adaptive immune response.

Acknowledgments
We thank P. Burrows, M. Cooper, J. Kearney, and Rolf F. Maier for their invaluable advice and support.

Disclosures
The authors have no financial conflicts of interest.

References
diversity gene loci, of which DFL16 and DSP2 originate from the same primordial Dα gene. 