B CELL DEVELOPMENT

VpreB serves as an invariant surrogate antigen for selecting immunoglobulin antigen-binding sites

Mohamed Khass,[1,2] Tessa Blackburn,[3] Peter D. Burrows,[3] Mark R. Walter,[3] Emidio Capriotti,[4,5] Harry W. Schroeder Jr.[1,3]*

Developmental checkpoints eliminate B cells that synthesize defective immunoglobulin (Ig) heavy (HC) and light (LC) chains. The first checkpoint tests μHCs paired with VpreB/λ5 in a pre-B cell receptor (pre-BCR) to determine whether the μHC will be able to bind conventional LCs to form membrane IgM. VpreB and λ5 also create a sensing site that interacts with the μHC antigen-binding region complementarity-determining region (CDR)–H3; however, whether this site contributes to Ig repertoire selection and function is unknown. We analyzed the amino acid content of CDR-H3s from HCs cloned from living and apoptotic pre-B cells and from IgG-antigen structures. Using a panel of D_H gene–targeted mice, we showed that progressively reducing CDR-H3 tyrosine content increasingly impaired pre-BCR checkpoint passage. Counting from cysteine at framework 3 position 96, we found that VpreB particularly selected for tyrosine at CDR-H3 position 101 and that Y101 also bound antigen in IgG-antigen structures. Therefore, in addition to its stabilization role in the pre-BCR, VpreB also acts as an early invariant antigen by selecting for particular CDR-H3 amino acids. These interactions shape the specificity of the IgG humoral response and may thus impose limitations on development of certain neutralizing antibodies.

INTRODUCTION

The genes that encode immunoglobulins (Igs) are assembled in developing B cells by a series of gene segment rearrangement events that begin with variable (V), diversity (D_H), and joining (J) gene segments at the heavy chain (HC) locus to encode a μHC. Progeny cells then rearrange a κ or λ light chain (LC) gene, and the newly formed B cell expresses membrane IgM as an antigen receptor [B cell receptor (BCR)]. The process of B cell development is optimized to create a highly diverse antibody repertoire. The antigen-binding sites of the antibody are formed by the juxtaposition of six complementarity-determining regions (CDRs): three from the HC and three from the LC. V and J gene segments are locked into one reading frame (RF), but the D_H gene segments can rearrange into any one of six different RFs, during which two rounds of nontemplated N nucleotide addition can also occur. Thus, the inclusion of a D_H makes CDR-H3 the most variable component of the antigen-binding site, and it often plays a key role in antigen recognition.

To avoid production of defective HCs, developing B cells must pass through a series of quality control (QC) checkpoints that test the integrity and function of their Ig (1). The first checkpoint occurs during the transition from the early (Hardy fraction C) to late (Hardy fraction D) pre-B cell stage (2) and tests for the ability of a nascent μHC to associate with surrogate LC (SLC) to form a pre-BCR (1). The SLC consists of two noncovalently associated proteins: the V_L homolog VpreB and the $J_L C_L$ homolog λ5 (1). Conventional V_L contains conserved framework region 2 (FR2) amino acids that associate with HCs FR2 and FR4 (fig. S1) (1, 3) to form a supportive scaffold for the HCDRs. VpreB shares several of these amino acids with V_L; thus, the ability of the μHC to form a pre-BCR predicts that it will ultimately

be able to form an IgM BCR. Pre-B cells that fail to form a functional pre-BCR perish by apoptosis in the bone marrow (BM). Unlike conventional LCs, VpreB and λ5 are invariant, making the pre-BCR checkpoint quite stringent.

In addition to FR2 and FR4, VpreB associates with CDR-H3. The VpreB portion of the CDR-H3 sensing site (CDR-H3 SS) contains a set of charged or hydrophilic residues, three of which are conserved between human and mouse (R51, D57, and R101). These residues are rare or absent in conventional V_L (fig. S1) (4). Here, we sought to test whether the SLC could use VpreB as an invariant surrogate antigen to select for, and hence against, CDR-H3s with a certain content of amino acids (5). If so, such an early selection step would limit the preimmune antibody repertoire and could help explain why some Ig antigen-binding sites are rare and thus more difficult to elicit after vaccination or infection (6).

The global distribution of amino acids in the primary CDR-H3 repertoire is already known to be nonrandom and specifically enriched for tyrosine (7). In part, this bias reflects the nonrandom use of individual D_H RFs (8), each of which exhibits a characteristic amino acid signature. In the preimmune repertoire, RF1 is the most frequently used, RF2 and RF3 are less frequently used, and the inverted RFs (iRFs) are rarely used (9). RF1 is enriched for tyrosine (fig. S2) (7); RF2, RF3, iRF2, and iRF3 are enriched for hydrophobic amino acids; and iRF1 is enriched for charged amino acids. RF3 tends to contain termination codons, limiting its use and thus the presence of leucine in CDR-H3.

The reduced use of RF2-encoded amino acids in mice has been attributed to the presence of an upstream ATG codon in frame with RF2 as a translation start site (8), which results in the production of a truncated protein product of DJ rearrangement termed Dμ. Although lacking the V_H component of the HC, Dμ can associate with the SLC to create a "sterile" pre-BCR that prevents further development. However, this blockade is not absolute, and one-fifth of in-frame VDJ rearrangements in early pre-B cells use RF2 (10, 11).

Four RF1 CDR-H3 tyrosines (Y101 to Y104), each of which can interact with CDR-H3 SS amino acids by means of hydrogen bonding, potentially stabilizing the complex, are observed in the only published pre-BCR crystal structure [Protein Data Bank (PDB): 2H32] (4). Because

[1]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35233, USA. [2]Division of Genetic Engineering, National Research Center of Egypt, Cairo, Egypt. [3]Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA. [4]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL 35233, USA. [5]Institute for Mathematical Modeling of Biological Systems, Department of Biology, University of Düsseldorf, Düsseldorf, Germany.
*Corresponding author. Email: hschroeder@uabmc.edu

RF2 does not encode such tyrosines, we postulated that the halving of RF2 usage that marks passage through the pre-BCR checkpoint (10, 11) might reflect reduced stability of pre-BCRs containing RF2-encoded amino acids in CDR-H3. Support for this view comes from studies in human where Dμ protein is not made, yet RF2 is still underrepresented in CDR-H3 (7), and from enrichment for CDR-H3s with RF2-encoded amino acids among Igs from a λ5, and thus pre-BCR–deficient, patient (12).

To test whether amino acids encoded by RF2 are subject to selection by the pre-BCR checkpoint, we took advantage of our panel of BALB/c mice with homozygous gene-targeted D_H alleles (13). The ΔD-DFL D_H allele contains a single wild-type (WT) DFL16.1 gene segment, which is the most frequently used D_H. The ΔD-DμFS D_H allele contains a single frameshifted DFL16.1 that favors the use of RF2 because the ATG Dμ start codon is in frame with RF1 and RF3 still contains a stop codon (fig. S2). Twenty percent of mature B cells in WT mice use RF2, whereas 30% of ΔD-DFL and 60% of ΔD-DμFS B cells use this RF (14). Thus, B cells in our panel create Ig CDR-H3s with a progressive increase in use of RF2 (13), which allowed us to test whether increased use of RF2-encoded amino acids influenced successful passage through the pre-BCR checkpoint.

RESULTS

Increasing the use of RF2-encoded amino acids reduces pre-BCR expression

We used the scheme of Hardy and Hayakawa (2) to identify progenitor B cell subsets in the BM. Fraction B contains pro-B cells that express SLC proteins and are undergoing HC rearrangement; fraction C contains early pre-B cells that express μHC and SLC; fraction C′ contains larger early pre-B cells where successful expression of pre-BCR is associated with cell division; and fraction D contains late pre-B cells that no longer express SLC and are undergoing LC rearrangement.

We examined pre-BCR formation, cell cycling, and apoptosis in fractions B to D from WT, ΔD-DFL, ΔD-DμFS, and λ5-deficient (λ5KO) mice, using the latter as a control for the absence of a pre-BCR. Among these mice, increased incorporation of RF2-encoded amino acids was associated with diminished pre-BCR formation, reduced cell division, and enhanced apoptosis, each of which is a hallmark of failure to pass the pre-BCR QC checkpoint.

To assess pre-BCR expression, we used μHC-specific polyclonal antibodies and a monoclonal antibody (mAb) that only recognizes a completely formed pre-BCR and not its separate components (15). Because cell surface pre-BCR expression is limited, intracellular staining techniques were used (fig. S3). When compared to WT mice (n = 8), there was a major decrease in the frequency of pre-BCR–expressing pre-B cells among ΔD-DμFS fractions C and C′ (P = 0.003 and P < 0.0001, t test; n = 8) (Fig. 1A). The ΔD-DFL pre-B cells were intermediate between WT and ΔD-DμFS (P = 0.03 and P = 0.0003, t test; n = 8) in fractions C and C′, respectively, which correlated with the more frequent use of RF2 with this gene segment when compared to the other D_H (14).

Increasing the use of RF2-encoded amino acids decreases pre-B cell proliferation

When compared to WT littermate controls, we previously found that although ΔD-DFL and ΔD-DμFS mice had normal absolute numbers of early pre-B cells [fraction C (C + C′)], there was a 17 and 40% reduction in the absolute number of late pre-B cells, respectively (13). To assess cell division, we evaluated bromodeoxyuridine (BrdU) incorporation after intraperitoneal injection (fig. S4). Increased use of RF2 correlated with a decrease in BrdU incorporation (fig. S5A). In all four Hardy fractions studied, WT mice incorporated the most BrdU. ΔD-DFL cells were similar to WT or intermediate between WT and ΔD-DμFS, depending on the fraction analyzed. Cell division was significantly reduced among ΔD-DμFS B lineage cells. BrdU incorporation among fractions B, C, and C′ cells was similar to that observed in λ5-deficient cells. The BrdU⁺ cells in fraction D likely represent pre-B cells that proliferated when they were passing through fraction C′ because pre-B cells can progress through the small pre-B cell stage (from large cycling pre-B cells to immature B cells) in approximately 5 hours and the mice in our studies were exposed to BrdU for 12 hours (16).

To refine the analysis of cell division, we used 7–amino-actinomycin D (7AAD) and BrdU staining to examine cell cycling (fig. S6). The hierarchy of cells in S phase (Fig. 1B) correlated with total BrdU incorporation (fig. S5A); more WT than ΔD-DμFS or λ5KO pre-B cells were replicating. The average frequency of replicating ΔD-DFL pre-B cells tracked with WT but individual mice demonstrated a greater variance than those from the other three strains. The hierarchy of G0 quiescent cells was complementary. ΔD-DμFS and λ5KO pre-B cells were more quiescent than WT, and ΔD-DFL pre-B cells were either similar to WT or intermediate between WT and ΔD-DμFS (fig. S5B).

Increasing the use of RF2-encoded amino acids increases pre-B cell apoptosis

Because apoptosis is a mark of failure to pass the pre-BCR QC checkpoint, we used staining with annexin V to quantify this process in the different mouse strains. Among pro-B cells, which have not yet reached the pre-BCR checkpoint, the percentage of annexin V⁺ cells was similar (Fig. 1C). However, among both early and late pre-B cells, more ΔD-DμFS than WT cells were undergoing apoptosis. For ΔD-DFL, apoptosis was similar to WT in early pre-B cells and then increased among late pre-B cells, lying intermediate between WT and ΔD-DμFS cells. This suggests that the use of RF2-encoded amino acids in CDR-H3 is disfavored for pre-BCR formation and stability.

Mice forced to preferentially rearrange into RF2 select for CDR-H3 Y101

A key prediction of our hypothesis is that passage through the pre-BCR checkpoint would be associated with changes in the prevalence of particular amino acids within CDR-H3. CDR-H3 can be divided into a base and a loop, the latter encoded by the D, by N nucleotides, and by the N terminus of the J. It rests on the surface of the V domain and can assume multiple conformations. The C terminus of the loop is enriched for J-encoded amino acids and exhibits less sequence diversity. Thus, we focused on the amino acids in the N-terminal end of CDR-H3, which is primarily encoded by N addition and D_H.

Counting from cysteine at position 96 (PDB numbering), we examined amino acids 99 to 103 among a large database of previously published sequences from WT and D-altered mice (file S1) (13). Although CDR-H3 is of variable length, position 101 tends to be centrally located in mouse Igs. Approximately $^2/_3$ of the amino acids at 101 were D-derived, $^1/_4$ were created by N addition, and $^1/_{10}$ were J-encoded.

To assess the contribution of individual amino acids at positions 99 to 103 before and after pre-BCR QC selection, we subtracted the fraction of usage per amino acid in fraction B (pro-B) from fraction D (late pre-B) cells (Fig. 2A). No major fractional changes in positional amino acid distribution were observed among WT (Fig. 2A). Thus, CDR-H3
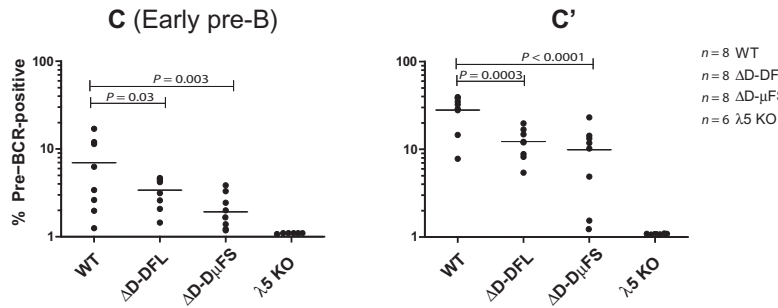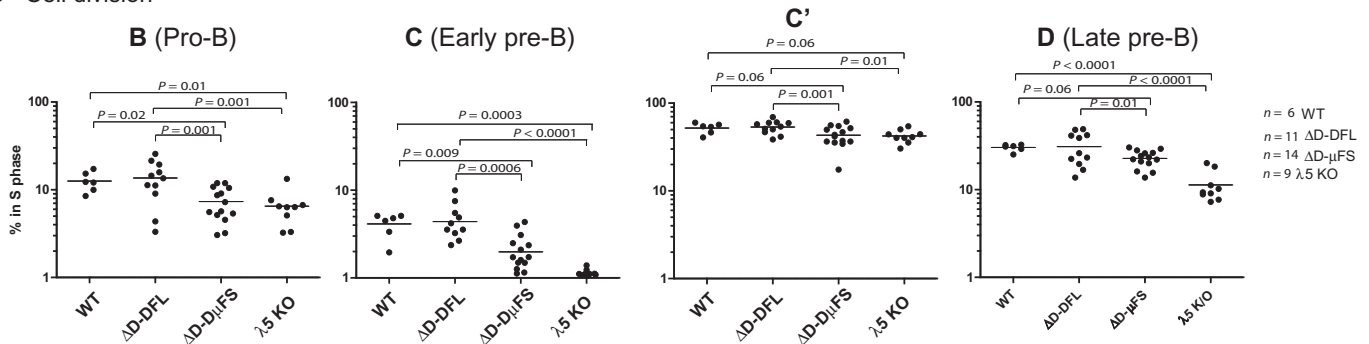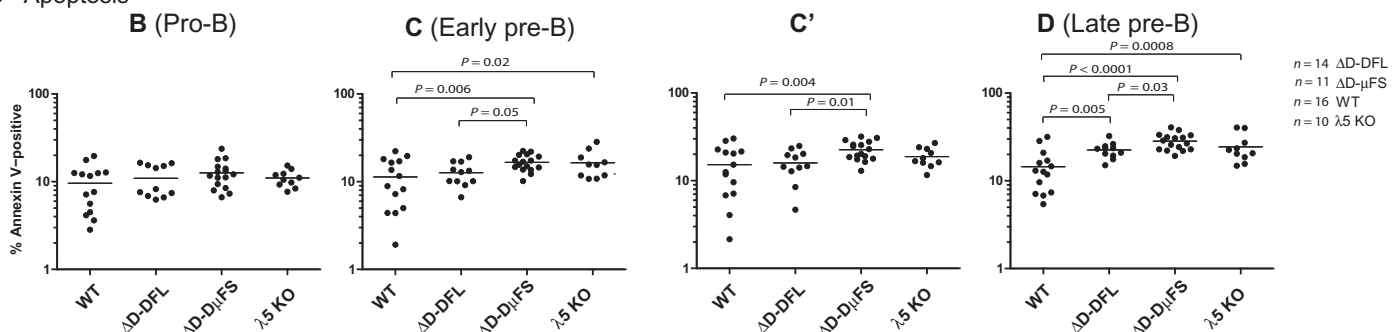
Fig. 1. Increasing the use of RF2-encoded amino acids in CDR-H3 diminishes pre-BCR formation, decreases cell cycling, and enhances apoptosis. (**A**) Among progenitor B lineage cells: percent of pre-BCR–positive (pre-BCR⁺) cells. The cells were intracellularly stained with μHC-specific polyclonal antibodies and with the SL156 monoclonal pre-BCR–specific antibody that detects a conformational epitope on the complete pre-BCR. Cells stained with both antibodies were counted as pre-BCR⁺. (**B**) Percent of cells in S phase as determined by BrdU incorporation. (**C**) Percent of cells undergoing apoptosis as determined by annexin V staining. Each dot indicates a measurement from an individual mouse. Dots are grouped by mouse strain (WT, ΔD-DFL, ΔD-DμFS, and λ5) and B cell developmental stage.

amino acid prevalence in WT B cells passing though the pre-BCR QC (Fig. 2B) appeared to maintain the amino acid preferences initially created by preferential use of D$_H$ RF1 (fig. S2).

Unlike WT, in ΔD-DμFS, a number of striking differences were observed after pre-BCR QC (Fig. 2, C and D). These included selection for tyrosine (RF1) at 101 and 102 and against threonine (RF2) and leucine (RF3) at 102. We also observed a decreasing fraction of phenylalanine in positions 101 and 102. This suggested that pre-BCR selection can be influenced by specific amino acid usage at particular CDR-H3 positions.

RF2- and RF3-encoded hydrophobic amino acids at CDR-H3 position 101 predispose to apoptosis

To better reveal the CDR-H3 molecular determinants for pre-BCR selection in WT, we pooled BM from twenty 8-week-old BALB/c mice;

stained cells from Hardy fractions C, C′, and D with annexin V and propidium iodide (PI) (fig. S7A); isolated the double-positive, late apoptotic cell fraction; and used reverse transcription polymerase chain reaction (RT-PCR) to clone and sequence CDR-H3s in cells that had failed pre-BCR QC (file S2). We compared this set of CDR-H3s to those previously isolated from live cells of the same fractions (17) and found that apoptotic cells were more likely to use amino acids encoded by RF2 or RF3 (fig. S7B).

A comparison of CDR-H3 sequences in living and apoptotic fraction C early pre-B cells revealed that apoptotic cells exhibited increased use of glycine (RF1) at position 100. The presence of glycine precludes the formation of side-chain interactions. Living cells demonstrated increased use of tyrosine and glycine (RF1) at 101, decreased use of leucine (RF3) at 101, and decreased threonine (RF2) at 102 and 103 (Fig. 3, A and D). Tyrosine provides a bulky, flexible
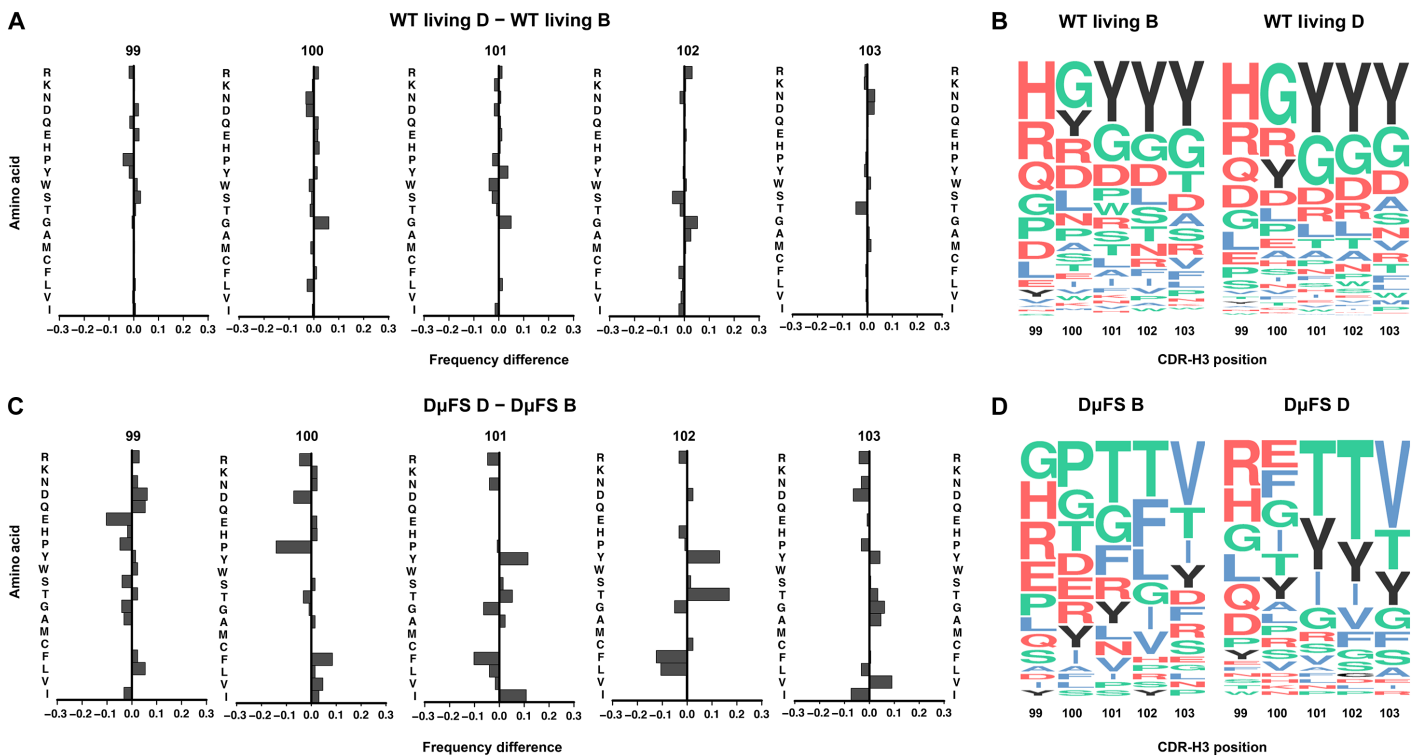
Fig. 2. CDR-H3 amino acid usage by position in pro-B (fraction B) versus late pre-B (fraction D) cells. (A and C) Fractional difference in the percent usage of individual amino acids between WT (A) and ΔD-DμFS (C) CDR-H3s at positions 99 to 103. Amino acids are arranged by hydrophobicity in the order of R, K, N, D, Q, E, H, P, Y, W, S, T, G, A, M, C, F, L, V, and I. **(B and D)** CDR-H3 amino acid usage by percentage rank order. The list of CDR-H3 sequences from WT and ΔD-DμFS cells in fractions B and D is reported in file S1.

side chain that permits changes in conformation and a terminal hydroxyl group that permits hydrogen bonding. Threonine has a hydroxyl group but lacks the bulky side chain. Leucine is hydrophobic and does not engage in hydrogen bonding with side-chain atoms. When compared to fraction C′ apoptotic cells, living fraction C′ cells exhibited an increase in tyrosine (RF1) and a decrease in glycine (RF1) and leucine (RF3) at position 100, and an increase in tyrosine (RF1) alone at position 101 (Fig. 3, B and D). These differences were magnified when compared to fraction C apoptotic cells (Fig. 3, C and D). These findings supported the hypothesis that amino acids encoded by RF2 and RF3 were disfavored and amino acids encoded by RF1, especially tyrosine, were favored for successful pre-BCR formation.

CDR-H3 Y101 has a beneficial effect on passage through the pre-BCR QC checkpoint

To gain further insight into the mechanism by which highly prevalent amino acids at specific CDR-H3 positions might be selected, we examined the crystal structure of the pre-B receptor (PDB: 2H32, Fig. 4A). This is a human pre-BCR structure and contains serine at position 59. We modeled the murine pre-BCR CDR-H3 SS by replacing human S59 with murine H59. CDR-H3 in this structure contains tyrosines at positions 100 to 103, all of which interact with the VpreB CDR-H3 SS. Of these, Y101 forms the most extensive hydrogen bond network and van der Waals contacts, with its hydroxyl group positioned for interactions with VpreB R51, D57, and R101. The side chains of the three conserved SLC CDR-H3 SS residues (fig. S1) form a pocket that can bind tyrosine at CDR-H3 position 101 (Fig. 4A), and the main-chain hydroxyl group of VpreB R101 can form a hydrogen

bond with CDR-H3 Y103. Finally, analysis of the original human pre-BCR complex and the modeled murine CDR-H3 SS predicts that CDR-H3 Y102 may form a hydrogen bond with mouse VpreB H59.

Our experimental data support a role for invariant VpreB in selecting for μHCs that contain a tyrosine at position 101. To test the structural and functional consequences of this selection, we generated and then analyzed molecular models of pre-BCR structures that used CDR-H3s cloned from the living and apoptotic pre-B cells (Fig. 4, B to D). We estimated the contribution of tyrosine at CDR-H3 positions 101, 102, and 103 to pre-BCR formation in sequences from living C and C′ pre-B cells that either lacked or contained tyrosine (Fig. 4C, top panel). We compared the probability of interaction (fraction of models with at least one hydrogen bond) with the VpreB CDR-H3 SS (see the Supplementary Materials). To avoid the potential confounding factor of runs of tyrosines, we focused this site-specific analysis on those sequences with tyrosine at the position of interest that lacked flanking tyrosines (1 left, 1 right). At positions 101 and 103, the distribution of the probability of interaction with the CDR-H3 SS when tyrosine is present (black) or not (white) achieved significance ($P < 0.01$ and $P = 0.03$, respectively, Kolmogorov-Smirnov test; table S1) (Fig. 4C, top panel).

We then grouped the sequences from WT living and apoptotic C and C′ pre-B cells into those with either <5% or ≥5% of the models predicting one or more hydrogen bonds between amino acids 101 to 103 and the CDR-H3 SS (table S2). CDR-H3 position 103 showed the highest fraction of interacting sequences (Fig. 4C, top panel), possibly due to its ability to form a hydrogen bond with the main chain of VpreB R101 (Fig. 4A). However, the most significant differences
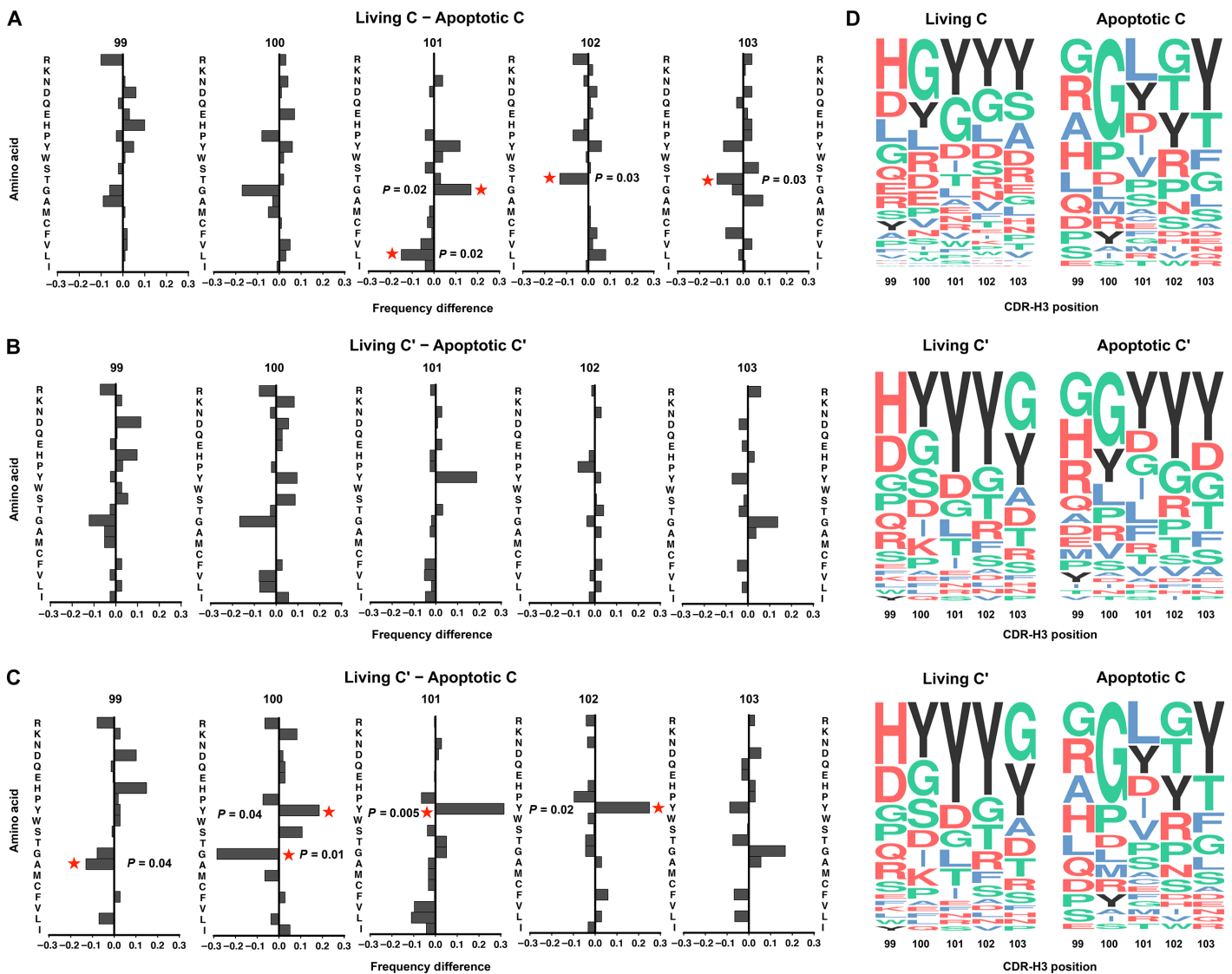
Fig. 3. CDR-H3 amino acid usage by position from WT living and apoptotic early pre-B cells. (**A** to **C**) Fractional difference in the percent usage of individual amino acids between CDR-H3 sequences cloned from apoptotic and living fraction C and C′ pre-B cells at positions 99 to 103. Amino acids are arranged by hydrophobicity as in Fig. 2. (**D**) CDR-H3 amino acid usage by percentage rank order. The list of CDR-H3 sequences from living and apoptotic cells in fractions C and C′ is reported in file S2.

between the modeled pre-BCR structures from apoptotic C and those from living C and C′ sequences were at position 101 (P = 0.04 and 0.01, respectively, one-tailed Fisher's exact test; table S3) (Fig. 4C, bottom panel). Thus, although the presence of a tyrosine at either position 101 or position 103 significantly increases the probability of forming hydrogen bonds with VpreB, the major differences between the apoptotic and living cells reflected the interaction of the CDR-H3 SS with CDR-H3 position 101.

The importance of position 101 is illustrated in a comparison of the models of a living C′ sequence that uses the same V_H7183-10 gene segment as an apoptotic C sequence (Fig. 4D). The CDR-H3s from these sequences differ in length by only one amino acid. When superimposed, their structures show a root mean square deviation of only 1.5 Å for the modeled structures as a whole and 3.5 Å for the CDR-H3 region (fig. S8). CDR-H3 Y101 in the modeled living cell sequence is adjacent to VpreB R51, D57, and R101, and Y102 is adjacent to VpreB H59 (Fig. 4, B and D). Conversely, CDR-H3 L101 in the

modeled apoptotic cell sequence is distant from VpreB, although CDR-H3 R102 is adjacent to VpreB R101 (Fig. 4D).

Further support for the beneficial effect of Y101 on passage through the pre-BCR QC derives from the distribution of amino acids at CDR-H3 101 in the sequences from living versus the apoptotic cells (Fig. 3). Y101 proved more common in living C′ sequences than in apoptotic C (P = 0.005, one-tailed Fisher's exact test; table S4) (Fig. 3C). Conversely, L101, which is frequently encoded by RF3, is more common in apoptotic C than in living C (P = 0.02, one-tailed Fisher's exact test; table S5) (Fig. 3A). Details regarding the comparative modeling protocol used for predicting the structures and their structural comparisons are reported in the Supplementary Materials.

In IgG Fab–antigen crystal structures, CDR-H3 Y101 often interacts with antigen

Having shown that CDR-H3 tyrosines, in particular Y101, are selected for and interact with VpreB during early B cell development, we next
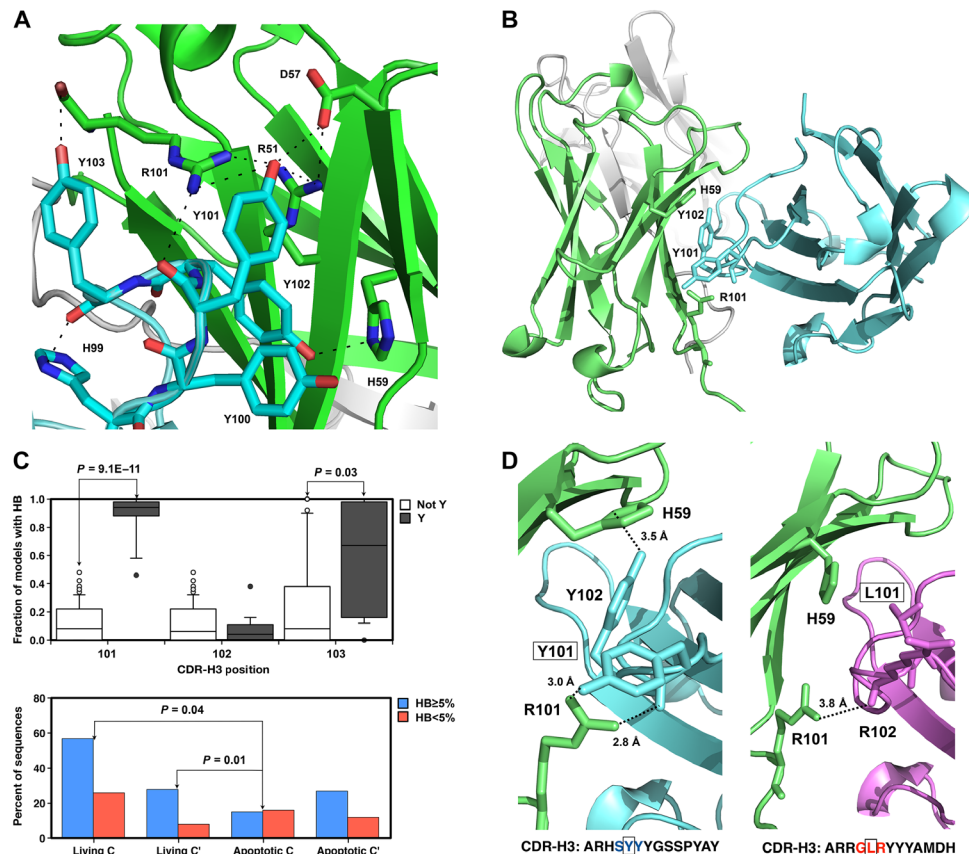
Fig. 4. Analysis of predicted pre-BCR complex structures. (**A**) Structure of the murine pre-BCR obtained by modeling the H59 side chain, which is a serine in the human VpreB structure (PDB: 2H32). VpreB is in green, and CDR-H3 residues 99 to 103 are in cyan. (**B**) Example of a predicted structure of a mouse pre-BCR complex from a WT living C' sequence. VpreB is in green, λ5 is in gray, and the HC V domain is in cyan. (**C**) Analysis of the pre-BCR models from CDR-H3 sequences in living C and C' cells. The top panel shows the fraction of interacting models for the subset of structures with (black, Y) and without (white, Not Y) tyrosine in the indicated CDR-H3 position for those sequences without a tyrosine flanking that position. In the bottom panel, we compared the percentage of sequences from living and apoptotic C/C' cells forming a hydrogen bond (HB) with CDR-H3 SS at position 101 only in at least 5% of the models (blue bar, HB ≥5%) and with sequences that engage in hydrogen bond with position 101 in less than 5% of the models (red bar, HB <5%). (**D**) Comparison of the predicted structures of two almost identical VDJCμ sequences from living C' and apoptotic C cells that use the same V_H7183-10 gene segment and differ only in CDR-H3. The predicted interactions between the SLC CDR-H3 SS (green) and the CDR-H3 region from a living C' cell (left, cyan) and an apoptotic C' cell (right, magenta).

investigated the role of these tyrosines in the interaction of the Fab with antigen using the IMGT (ImMunoGeneTics) database to identify a set of 133 nonredundant murine antigen–Fab crystal structures in the PDB (www.pdb.org) (file S3). A comparison of the distribution of amino acids at positions 99 to 103 revealed that G99 was preferred in the IgG Fabs rather than H99, which was more prevalent in the living progenitor B cells (Figs. 2B, 3D, and 5A). However, in positions 100 to 103, both the preimmune living progenitor B cells and the antigen-selected IgGs demonstrated the same preference for G100, Y101, Y102, and Y103 (Fig. 5A).

We examined the IgG-antigen crystal structures to determine which CDR-H3 residues used a hydrogen bond to engage with antigen (Fig. 5B) or LC. We found that the 101–103 tripeptide (101,102, and 103) was more likely to engage antigen (65% of cases) than LC (42%) (table S6). As above, we then considered only the structures with a tyrosine at 101, 102, or 103 that lacked flanking tyrosines. At positions 101 and 103, structures that use tyrosine (black bar) were significantly more likely to engage antigen with a hydrogen bond than those with residues other than tyrosine (white bar) ($P = 0.008$, one-tailed Fisher's exact test; table S7) (Fig. 5C). No significant preferences

for specific amino acid composition were found for structures binding LC, which is more likely to engage CDR-H3 at position 103 than position 101 (table S8).

DISCUSSION

μHCs in late pre-B cells have access to an array of diverse LCs. Thus, the only time during B cell development when the LC partner is invariant occurs at the pre-BCR QC checkpoint. Here, we provide evidence that this checkpoint selects not only for cells expressing μHC able to bind to a conventional LC but also for the use of particular amino acids at specific positions at the center of the antigen-binding site, where they can subsequently influence antigen recognition and antibody production. This selection process appears to begin just after the μHC is created, in the endoplasmic reticulum (ER) of the early pre-B cell. The presence of a CDR-H3 that cannot bind effectively to the CDR-H3 SS of the SLC would result in retention of the μHC by ER chaperones and failure to form a pre-BCR, resulting in decreased cell proliferation and increased cell loss by apoptosis.
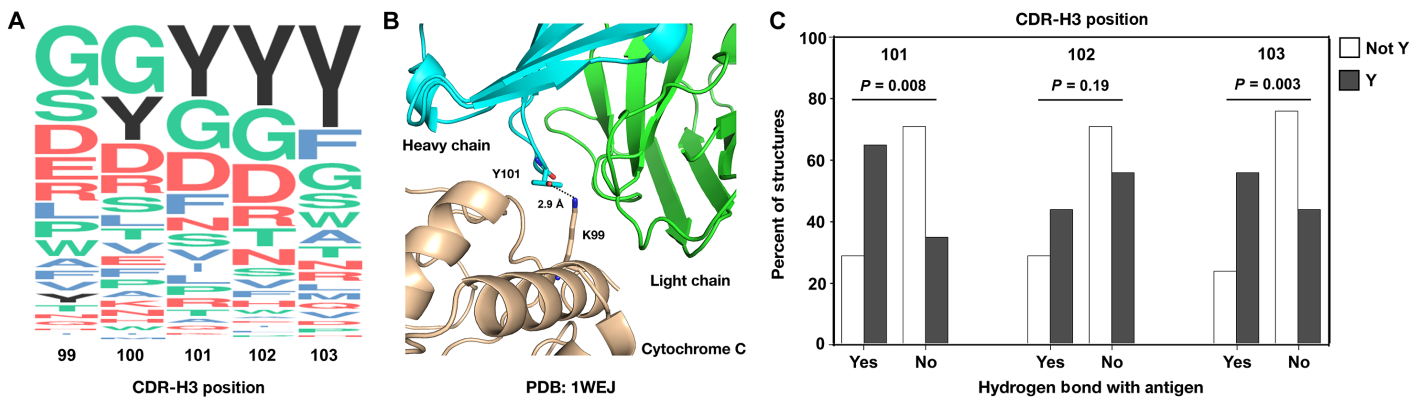
Fig. 5. Analysis of the CDR-H3 region in murine IgG Fab–antigen PDB structures. (**A**) Amino acid composition of CDR-H3 regions at positions 99 to 103 in the 133 analyzed structures. (**B**) Example of CDR-H3 Y101 interacting with antigen (cytochrome C) in PDB: 1WEJ. (**C**) Fraction of structures with (black, Y) and without (white, Not Y) tyrosine in CDR-H3 positions 101 to 103 that form hydrogen bonds with antigen among those sequences where the tyrosine in the selected position lacks a flanking (1 left, 1 right) tyrosine. The list of Fab-antigen PDB structures used in this analysis is reported in file S3. The procedure used for the selection of the Fab-antigen PDB structures is described in the Supplementary Materials.

Although not identical, the level of sequence similarity (>70%) between human and mouse VpreB suggests that the selection mechanisms described in mouse may also be used in human to regulate CDR-H3 amino acid content and RF usage (*4*). We have shown that changing the pattern of D_H RF usage, and thus the prevalence of tyrosine in CDR-H3, in mice alters patterns of epitope recognition after vaccination with HIV gp140 (*6*). It is possible that the binding instability of SLC to μHCs containing CDR-H3s with amino acids commonly encoded by non–tyrosine-enriched RFs may help explain the difficulty experienced by HIV patients in generating neutralizing antibodies that require CDR-H3 amino acids typically encoded by RF2 or RF3.

We have previously shown that natural selection of germline D_H sequence creates a bias in global CDR-H3 amino acid content that is established at the time of VDJ rearrangement before antigen exposure (*17*). The presence of other tyrosines around position 101 and their possible interactions with VpreB support the hypothesis that the binding of CDR-H3 to the SLC is a stochastic process in which specific amino acids increase the chance of forming hydrogen bonds that stabilize the pre-BCR.

Although among individual structures the likelihood of interactions between the antigen and individual CDR-H3 residues can be influenced by the length of the CDR-H3 loop, our observations support the global hypothesis that CDR-H3 residues 101 and 103 often play an important role, especially in the presence of tyrosine, in binding both the VpreB in the pre-BCR (Fig. 4D) and the antigen in the IgG-antigen complex (Fig. 5C). High-throughput sequencing of HCs from developing pre-B cells from the WT and D-altered mice would enable enhanced comparisons of CDR-H3s of similar length.

Both natural selection of D_H germline sequence and passage through the pre-BCR checkpoint promote the inclusion of a tyrosine at CDR-H3 position 101. However, our results do not exclude alternative interaction patterns in which amino acids with similar chemico-physical properties can form stable pre-BCR. In our view, the CDR-H3 regions from apoptotic cells correspond to amino acid combinations with a low probability of forming hydrogen bonds with VpreB R51/D57/R101 component of the CDR-H3 sensing site and thus creating a functional pre-BCR with an HC antigen-binding site optimized for binding antigen. An analysis of repertoire selection and responses to antigenic challenges in mice, where sequence of

the CDR-H3 SS has been mutated, would provide confirmation of this interpretation.

In summary, the sequence of D_H has a positional effect, which selects for particular amino acids at specific locations within the antigen-binding site (Fig. 2) (*6*). Here, we show that selection of CDR-H3 by the VpreB portion of SLC reinforces the positional bias in amino acid content, which is consistent with the bias in antigen-binding IgG. Our findings suggest that the molecular mechanism underlying the selection of CDR-H3 Y101 is the stabilization of the pre-BCR by the creation of hydrogen bonds between Y101 and VpreB R51/D57/R101. Analysis of the crystal structures of IgG Fab–antigen complexes indicates that the CDR-H3 residues selected by VpreB are ultimately more likely to form hydrogen bonds with antigen than LC. The combination of natural selection of D_H sequence and somatic selection by invariant SLC favors the presence of tyrosine at key positions in the antigen-binding site, thus shaping the antigen specificity of the primary (IgM) and secondary (IgG) humoral response.

MATERIALS AND METHODS
Study design
We sought to determine whether pre-BCR expression, progenitor B cell division, and progenitor B cell apoptosis were influenced by the amino acid sequence of Ig CDR-H3 during B cell development by analyzing a panel of BALB/c mouse strains that had been gene-targeted in their D_H locus and thus expressed a gradient of altered amino acid sequence. We then performed a combined sequence structural analysis of HCs cloned from living and apoptotic progenitor B cells to test for positional patterns of amino acid preference and compared the results to IgG Fab–antigen structures that were publicly available.

All data obtained from the mice are presented, including outliers. We used six or more mice per group. The genotypes were not blinded, and randomization was not performed. All of the modeled HCs cloned from early pre-B cells were unique. To select a representative set of structures, we queried the IMGT database (http://www.imgt.org/) searching for *Mus musculus* Fab with protein antigen and resolution ≤3.5 Å. From this initial group of structures, we selected the PDBs with HC sequence similar to our set of V_H7183 family members. The sequence similarity was detected with the BLAST algorithm (*18*) using an *e*-value threshold of 10^{-3}. Redundant PDBs were removed,

selecting only the representative structures of each cluster defined using a 95% sequence similarity threshold (ftp://resources.rcsb.org/sequence/clusters/clusters95.txt). The final set of 133 Fab-antigen PDB structures is provided in file S3.

Mice

All experiments and animal procedures were performed using protocols approved by the University of Alabama at Birmingham (UAB) Institutional Animal Care and Use Committee. BALB/c mice were bred under specific pathogen–free conditions in the UAB vivarium. The ΔD-DFL and ΔD-DμFS D-altered mice were generated on a BALB/c background as described previously (14, 19). λ5KO mice, originally on a C57BL/6 background (20), were backcrossed for 22 generations onto BALB/c.

Flow cytometry analysis of BM B cell subsets

Single-cell suspensions were prepared from femurs of 8-week-old mice as previously described (7). Cells were washed, stained, and analyzed on an LSRII (BD Biosciences) using the following labeled antibodies: anti-B220 (Pacific Blue), anti-CD19 [allophycocyanin (APC)–Cy7], anti-CD43 (biotin), and anti–BP-1 [phycoerythrin (PE)] from BD Pharmingen and anti-AA4.1 (PE-Cy7) from eBioscience. PI was used to identify dead cells. The scheme of Hardy and Hayakawa (2) was used to identify different B cell subsets in the BM. Fractions C and C′ were distinguished on the basis of cell size.

Intracellular evaluation of binding of μHC to SLC

Pre-B cell subsets in the BM, as identified by the expression of relevant surface markers, were fixed using a Cytofix/Cytoperm kit (BD Pharmingen) and then stained for intracellular pre-BCR using a pre-BCR mAb (clone SL156) (BD Pharmingen) and goat anti-mouse μHC [fluorescein isothiocyanate (FITC)] (Southern Biotech) (Fig. 1A). Successful formation of pre-BCR was defined by dual staining with both antibodies, as described by Kawano et al. (15) (fig. S3).

BrdU labeling

BrdU incorporation was used to assess B cell proliferation. BrdU was administered by intraperitoneal injection (1 mg per mouse) twice daily at 12-hour intervals. Twelve hours after the last injection, mice were killed, BM cells were washed and fixed, and BrdU-positive B cells were detected with antibodies specific for BrdU using the BrdU Flow Kit (BD Biosciences) according to the manufacturer's protocol (fig. S4).

Cell cycle analysis

BM B cell subsets were analyzed for cell cycle progression using BrdU and 7AAD staining according to the BrdU Flow Kit (BD Biosciences). Cell cycle stages were identified as follows: G0 phase, quiescent cells (BrdU-negative and 7AAD-negative); G1 phase, cells preparing for cycling (BrdU-positive and 7AAD-negative); S phase, proliferating cells (BrdU-positive and 7AAD-positive); and G2-M phase, cells in the process of finishing cycling (BrdU-intermediate and 7AAD-positive) (Fig. 1B, figs. S5 and S6).

Evaluating the extent of apoptosis in different B cell subsets

The frequency of apoptosis in different B cell subsets in the BM was determined on the basis of PI and annexin V staining using the FITC Annexin V Apoptosis Detection Kit II (BD Biosciences). Late apoptotic cells that express both annexin V and PI were sorted for CDR-H3 complementary DNA (cDNA) construction, sequencing (as described below), and structural modeling because they represent a true indicator of fully apoptotic cells (Fig. 1C and fig. S7).

Flow cytometric sorting, RNA preparation, RT-PCR, and sequencing

Living and late apoptotic pre-B cells were sorted on the basis of expression of surface markers (2) using the FACSAria cell sorter (BD Biosciences). Total RNA was isolated and converted to cDNA using V$_H$7183-specific VDJCμ RT-PCR primers. Clones positive for an insert of the appropriate size were sequenced and analyzed as previously described (14, 17, 21). A list of the unique, in-frame sequences used for analysis in this work is provided in file S2.

Data set for structural modeling

Total living and apoptotic cells were collected without bias. Sequence/structure-function relationships were determined using potential differences in the predicted structures obtained from living and apoptotic cells. We examined 31 and 39 unique sequences from apoptotic C and C′, and 83 and 36 sequences from living C and C′, respectively. The 189 CDR-H3 sequences are reported in file S2.

Sequence analysis of the CDR-H3 regions

CDR-H3 was defined as the sequence starting immediately after the cysteine at the end of FR3 at position 96 and then through, but not including, the tryptophan that begins FR4. At positions 99 to 103, we calculated the frequency of the amino acids in individual developing B cell subsets. The residues are grouped in three main categories according to their hydrophobicity as assessed in the Kyte-Doolittle scale (22) as modified by Eisenberg (23). Under this scheme, arginine, lysine, asparagine, aspartic acid, glutamine, glutamic acid, and histidine are considered charged and shown in red; proline, tryptophan, serine, threonine, and glycine are considered neutral or hydrophilic and shown in green; and alanine, methionine, cysteine, phenylalanine, leucine, valine, and isoleucine are considered hydrophobic and shown in blue. Tyrosine, which is the most abundant residue in our data set (~21%), is normally included among the hydrophilic, neutral fraction. For this work, tyrosine was counted separately and is shown in black for emphasis. The amino acid composition in the different positions was displayed using WebLogo (http://weblogo.berkeley.edu/) (24).

Structural analysis of the predicted pre-BCR models

To study the structural and functional features of the CDR-H3 region, we predicted the structure of the mouse pre-BCR using homology modeling (25) using the structure of the human pre-BCR (PDB: 2H32) as a template (4). The structure of this protein complex is a trimer containing VpreB, λ5, and Ig μHC.

All the murine pre-BCR models were then generated using the sequences of the mouse Vpre1 (SwissProt: VPRE1_MOUSE) and λ5 (SwissProt: IGLL1_MOUSE). Each model of the pre-BCR complex differs in the sequence of the Ig HC, which is assembled using the different V$_H$7183 family members with unique CDR-H3s. Global alignments were calculated with align (26) using default parameters without end-gap penalties. The sequence identities between human and mouse sequences are 77% for VpreB, 65% for λ5, and in the range of 40 to 58% for μHC depending on the specific V$_H$7183 family member (see alignments in file S4).

Mouse pre-B protein structures were predicted using MODELLER (27). We generated 50 models for each sequence in our data set. These were used to predict the interactions between the CDR-H3 regions,

VpreB and λ5 in the pre-BCR complex. An example of a comparison between two similar predicted structures of μHC sequences from living C′ (cyan) and apoptotic C (magenta) cells is shown in Fig. 4 (B and D) and fig. S8. We analyzed the region surrounding CDR-H3 residues 101 to 103 and their possible interaction with VpreB residues. A hydrogen bond was predicted considering a distance threshold of 3.5 Å between two electronegative atoms (N, O). For each CDR-H3 sequence, we calculated the probability of interacting models in a given position as the fraction of models in which a hydrogen bond was detected. To evaluate the importance of tyrosine in residues 101 to 103 and to reduce the bias of consecutive tryrosines, we evaluated a subset of sequences with and without tyrosine in the central position without flanking tyrosines. The analysis of this subset of sequences from living C and C′ cells is reported in the top panel of Fig. 4C. We then focused only on position 101 and considered the fraction of sequences forming a hydrogen bond (HB) with CDR-H3 SS in at least 5% of the models (blue bar, HB ≥5%) compared with the fraction of sequences that engaged in hydrogen bond in less than 5% of the models (red bar, HB <5%) across living and apoptotic fraction C and C′ sequences (Fig. 4C, bottom panel). The significance in the difference of the predicted hydrogen bonds in position 101 was estimated by one-tailed Fisher's exact test. All the PDB files of the predicted models are available at http://biofold.org/emidio/data/cdr-h3/models.tar.gz.

Structural analysis of the Fab-antigen PDB structures

Finally, we performed an analysis of murine Fab–antigen PDB structures. The structures were analyzed by calculating the hydrogen bonds using a distance threshold of 3.5 Å between two electronegative atoms (N, O). To estimate the importance of tyrosine in the formation of hydrogen bonds in positions 101 to 103, we compared the fraction interacting and noninteracting structures with and without tyrosine in a given position. The bias for consecutive tyrosines was reduced considering the subset of sequences without flanking tyrosines around the central position. The result of this analysis is reported in Fig. 5C.

Statistical analyses

Differences between populations or groups were assessed by one-tailed Student's t test, two-tailed Fisher's exact test, the Levene test for the homogeneity of variance, the Kolmogorov-Smirnov test, or the nonparametric Mann-Whitney test. A P value of less than 0.05 was considered significant. JMP version 12 (SAS Institute Inc.) was used for analysis.

SUPPLEMENTARY MATERIALS

immunology.sciencemag.org/cgi/content/full/1/1/aaf6628/DC1

Fig. S1. Comparison of critical residues in mouse and human SLC and LC.
Fig. S2. Percentage rank order of individual amino acids by position and by RF in the germline sequence of the BALB/c WT D_H repertoire and in the D-altered ΔD-DFL and ΔD-DμFS mice.
Fig. S3. Detection of pre-B cells expressing the pre-BCR.
Fig. S4. Quantification of BrdU incorporation in developing B cells by flow cytometry.
Fig. S5. Distribution of developing B cells that have incorporated BrdU or remained quiescent in G0.
Fig. S6. Flow cytometric gates used to evaluate the cell cycle distribution of developing BM B cell subsets.
Fig. S7. D_H RF usage among in-frame HCs cloned from living and apoptotic pre-B cells.
Fig. S8. Comparison of predicted structures of pre-BCRs from apoptotic C and living C′ cells.
Table S1. Statistical tests reported in the upper panel of Fig. 4C.
Table S2. Analysis of predicted hydrogen bond interactions in positions 101 to 103 for living and apoptotic cell fractions C and C′.
Table S3. Statistical tests reported in the bottom panel of Fig. 4C.
Table S4. Statistical tests reported in Fig. 3C.
Table S5. Statistical tests reported in Fig. 3A.
Table S6. Hydrogen bond interaction in the IgG Fab–antigen PDB structures in positions 101 to 103.

Table S7. Statistical tests reported in Fig. 5C.
Table S8. Hydrogen bond interaction in the IgG Fab–antigen PDB structures in positions 101 to 103* without flanking tyrosines.
File S1. List of the CDR-H3 sequences from WT and ΔD-DμFS cells in fractions B and D used in this work.
File S2. List of the V_H and CDR-H3 sequences from apoptotic and living C and C′ fractions used in this work.
File S3. Table of Fab-Ag crystal structures.
File S4. Sequence of the V_H regions and alignments used in the comparative modeling protocol.
File S5. Excel file containing tabulated data for Fig. 1 and fig. S5.
File S6. PDF file containing the results of the statistical analysis of the data for Fig. 1 and fig. S5.

REFERENCES AND NOTES

1. F. Melchers, Checkpoints that control B cell development. *J. Clin. Invest.* **125**, 2203–2210 (2015).
2. R. R. Hardy, K. Hayakawa, B cell development pathways. *Annu. Rev. Immunol.* **19**, 595–621 (2001).
3. J. E. T. Narciso, I. D. C. Uy, A. B. Cabang, J. F. C. Chavez, J. L. B. Pablo, G. P. Padilla-Concepcion, E. A. Padlan, Analysis of the antibody structure based on high-resolution crystallographic studies. *Nat. Biotechnol.* **28**, 435–447 (2011).
4. A. J. Bankovich, S. Raunser, Z. S. Juo, T. Walz, M. M. Davis, K. C. Garcia, Structural insight into pre-B cell receptor function. *Science* **316**, 291–294 (2007).
5. S. T. Kim, H. Shirai, N. Nakajima, J. Higo, H. Nakamura, Enhanced conformational diversity search of CDR-H3 in antibodies: Role of the first CDR-H3 residue. *Proteins* **37**, 683–696 (1999).
6. Y. Wang, P. Kapoor, R. Parks, A. Silva-Sanchez, S. M. Alam, L. Verkoczy, H. X. Liao, Y. Zhuang, P. D. Burrows, M. Levinson, A. Elgavish, X. Cui, B. F. Haynes, H. W. Schroeder Jr., HIV-1 gp140 epitope recognition is influenced by immunoglobulin D_H gene segment sequence. *Immunogenetics* **68**, 145–155 (2016).
7. I. I. Ivanov, J. M. Link, G. C. Ippolito, H. W. Schroeder Jr., in *The Antibodies*, M. Zanetti, J. D. Capra, Eds. (Taylor and Francis Group, 2002), vol. 7, pp. 43–67.
8. H. Gu, D. Kitamura, K. Rajewsky, D_H reading frame bias: Evolutionary selection, antigen selection or both? Evolutionary selection. *Immunol. Today* **12**, 420–421 (1991).
9. A. Silva-Sanchez, C. R. Liu, A. M. Vale, M. Khass, P. Kapoor, A. Elgavish, I. I. Ivanov, G. C. Ippolito, R. L. Schelonka T. R. Schoeb, P. D. Burrows, H. W. Schroeder Jr., Violation of an evolutionarily conserved immunoglobulin diversity gene sequence preference promotes production of dsDNA-specific IgG antibodies. *PLOS One* **10**, e0118171 (2015).
10. M. Zemlin, R. L. Schelonka, G. C. Ippolito, C. Zemlin, Y. Zhuang, G. L. Gartland, L. Nitschke, J. Pelkonen, K. Rajewsky, H. W. Schroeder Jr., Regulation of repertoire development through genetic control of D_H reading frame preference. *J. Immunol.* **181**, 8416–8424 (2008).
11. M. Khass, K. Buckley, P. Kapoor, R. L. Schelonka, L. S. Watkins, Y. Zhuang, H. W. Schroeder Jr., Recirculating bone marrow B cells in C57BL/6 mice are more tolerant of highly hydrophobic and highly charged CDR-H3s than those in BALB/c mice. *Eur. J. Immunol.* **43**, 629–640 (2013).
12. Y. Minegishi, M. E. Conley, Negative selection at the pre-BCR checkpoint elicited by human μ heavy chains with unusual CDR3 regions. *Immunity* **14**, 631–641 (2001).
13. H. W. Schroeder Jr., M. Zemlin, M. Khass, H. H. Nguyen, R. L. Schelonka, Genetic control of D_H reading frame and its effect on B-cell development and antigen-specific antibody production. *Crit. Rev. Immunol.* **30**, 327–344 (2010).
14. R. L. Schelonka, I. I. Ivanov, D. H. Jung, G. C. Ippolito, L. Nitschke, Y. Zhuang, G. L. Gartland, J. Pelkonen, F. W. Alt, K. Rajewsky, H. W. Schroeder Jr., A single D_H gene segment creates its own unique CDR-H3 repertoire and is sufficient for B cell development and immune function. *J. Immunol.* **175**, 6624–6632 (2005).
15. Y. Kawano, S. Yoshikawa, Y. Minegishi, H. Karasuyama, Selection of stereotyped V_H81X-μH chains via pre-B cell receptor early in ontogeny and their conservation in adults by marginal zone B cells. *Int. Immunol.* **17**, 857–867 (2005).
16. R. Casellas, T.-A. Y. Shih, M. Kleinewietfeld, J. Rakonjac, D. Nemazee, K. Rajewsky, M. C. Nussenzweig, Contribution of receptor editing to the antibody repertoire. *Science* **291**, 1541–1544 (2001).
17. I. I. Ivanov, R. L. Schelonka, Y. Zhuang, G. L. Gartland, M. Zemlin, H. W. Schroeder Jr., Development of the expressed Ig CDR-H3 repertoire is marked by focusing of constraints in length, amino acid use, and charge that are first established in early B cell progenitors. *J. Immunol.* **174**, 7773–7780 (2005).
18. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
19. R. L. Schelonka, M. Zemlin, R. Kobayashi, G. C. Ippolito, Y. Zhuang, G. L. Gartland, A. Szalai, K. Fujihashi, K. Rajewsky, H. W. Schroeder Jr., Preferential use of D_H reading frame 2 alters B cell development and antigen-specific antibody production. *J. Immunol.* **181**, 8409–8415 (2008).

20. D. Kitamura, A. Kudo, S. Schaal, W. Müller, F. Melchers, K. Rajewsky, A critical role of λ5 protein in B cell development. *Cell* **69**, 823–831 (1992).

21. G. C. Ippolito, R. L. Schelonka, M. Zemlin, I. I. Ivanov, R. Kobayashi, C. Zemlin, G. L. Gartland, L. Nitschke, J. Pelkonen, K. Fujihashi, K. Rajewsky, H. W. Schroeder Jr., Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production. *J. Exp. Med.* **203**, 1567–1578 (2006).

22. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

23. D. Eisenberg, Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* **53**, 595–623 (1984).

24. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

25. M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Šali, Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).

26. X. Huang, W. Miller, A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337–357 (1991).

27. N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, A. Sali, Unit 5.6 Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **47**, 5.6.1–5.6.32 (2006).

Citation: M. Khass, T. Blackburn, P. D. Burrows, M. R. Walter, E. Capriotti, H. W. Schroeder Jr., VpreB serves as an invariant surrogate antigen for selecting immunoglobulin antigen-binding sites. *Sci. Immunol.* **1**, aaf6628 (2016).