Contribution of V<sub>H</sub> Gene Replacement to the Primary B Cell Repertoire

Zhixin Zhang,1,3 Michael Zemlin,1,3,6 Yui-Hsi Wang,1,3,5,11 Delicia Munfus,1,3,5 Leslie E. Huye,10 Harry W. Findley,1 S. Louis Bridges, Jr.,1,3,5 David B. Roth,10 Peter D. Burrows,1,9 and Max D. Cooper1,3,4,6,7,*

Introduction

A vast repertoire of antibody specificities is somatically generated in developing B cells in preparation for the challenge of non-self-antigens (Rajewsky, 1996). B lymphocytes generated with potentially harmful self specificities must be removed from this repertoire in order to reduce the possibility of autoimmunity (Goodnow et al., 1995). The diverse specificities are generated by recombination of previously separated variable (V), diversity (D) (for heavy chain only), and joining (J) gene segments to form the variable domain exons of immunoglobulin genes that are expressed by the B cell (Tonegawa, 1983; Bassing et al., 2002). This process is mediated by recombination-activating gene (RAG-1 and RAG-2) products that recognize the recombination signal sequences (RSS) flanking the V, D, and J gene segments (Schatz and Baltimore, 1988; Schatz et al., 1989; Oettinger et al., 1990). The consensus RSS is composed of a heptamer (CACTGTG) and a nonamer (GGTTTTTGT) separated by a nonconserved 12 or 23 base pair spacer region (Tonegawa, 1983; Lewis, 1994). The heptamer and nonamer sequences are important for RAG-1/2 binding and subsequent DNA cleavage and modifications of the conserved heptamer or nonamer nucleotides can therefore profoundly decrease recombination efficiency (Difilipantonio et al., 1996; Ramsden et al., 1996; Steen et al., 1996; Swanson and Desiderio, 1998; Kim and Oettinger, 1998). The spacer length between nonamer and heptamer also plays a crucial role in regulation of recombination. In what became known as the 12/23 rule, V(D)J recombination was found to occur preferentially between gene segments with 12 bp and 23 bp spacer RSS (Tonegawa, 1983; Lewis, 1994; van Gent et al., 1996; Eastman et al., 1996; Steen et al., 1997). During V(D)J rearrangement, the RAG protein complex first nicks the RSS at the end of each heptamer to allow formation of the coding end hairpin structure while leaving the signal ends as double-stranded DNA breaks; hairpin formation is the step controlled by the 12/23 rule (Gellert, 2002; Brandt and Roth, 2002). The two coding end hairpins are then opened (Ma et al., 2002) and joined to form the variable domain exons of immunoglobulin (for heavy chain only), and joining (J) gene segments (Schatz and Baltimore, 1988; Schatz et al., 1989; Oettinger et al., 1990). The consensus RSS is composed of a heptamer (CACTGTG) and a nonamer (GGTTTTTGT) separated by a nonconserved 12 or 23 base pair spacer region (Tonegawa, 1983; Lewis, 1994; van Gent et al., 1996; Eastman et al., 1996; Steen et al., 1997). The signal ends are also ligated and in the case of deletional recombination, the intervening DNA forms an extrachromosomal circular molecule (Gellert, 2002; Brandt and Roth, 2002).

Immunoglobulin gene rearrangement proceeds in a step-wise manner during early B lineage development. D<sub>H</sub>−→J<sub>H</sub> rearrangement occurs before the V<sub>H</sub>−→D<sub>H</sub> rearrangement and these gene rearrangements precede the V<sub>L</sub>−→J<sub>L</sub> rearrangement in the light chain gene loci (Burrows et al., 1979; Alt et al., 1984; Burrows and Cooper, 1997; Bassing et al., 2002). The expression of a functional κ heavy chain generated through VDJ recombination is a prerequisite for further development along the B lineage pathway (Rajewsky, 1996). Due to the imprecise processing of the coding ends, two-thirds of the V<sub>H</sub>−→D<sub>H</sub> joints and V<sub>L</sub>−→J<sub>L</sub> joints found during V(D)J rearrangement may be out of reading frame and thus unable to encode immunoglobulin chains. Cells with nonfunctional IgH or IgL gene rearrangements do not survive unless they are rescued by a subsequent functional rearrangement. Even after generating a V<sub>H</sub>D<sub>H</sub>κ open reading frame, the expressed κ heavy chains may fail to pair with surrogate light chains or with conventional light chains to form the functional pre-BCR or BCR needed to promote further differentiation (ten Boekel,...
et al., 1998). Finally, B cells that possess self-reactive antigen receptors require alteration of their antigen specificity or else must be eliminated before their release into the periphery. In order to be rescued, these early B lineage cells must retain the capability to edit the initially generated variable region exons (Radic and Zouali, 1996; Nussenzweig, 1998; King and Monroe, 2001).

The organization of the \( V_\lambda \) and \( J_\lambda \) gene segments within the Ig \( \kappa \) and \( \lambda \) loci allows secondary rearrangement to occur by joining of an upstream \( V_\lambda \) and a downstream \( J_\lambda \) gene segment because they are flanked by complementary RSS sites. The light chain gene secondary rearrangement is mechanistically similar to the primary rearrangement and can thus be performed as long as the locus remains accessible and the recombination machinery is still operative (Radic and Zouali, 1996). With each round of rearrangement, a new \( V_\lambda -J_\lambda \) joint is formed and the previous \( V_\lambda -J_\lambda \) joint is deleted (Casellas et al., 2001; King and Monroe, 2001). In addition, a cell with a nonfunctional \( \kappa \) rearrangement still has the option to rearrange \( \lambda \) light chain genes (Casellas et al., 2001; King and Monroe, 2001).

Secondary rearrangement of an upstream \( V_\lambda \) to a pre-formed \( D_\lambda J_\lambda \) gene provides conceptual difficulties, however, because deletion of the intervening D segments during the primary rearrangement to \( D_\lambda J_\lambda \) rearrangement leaves no 12 bp RSS to recombine with the \( V_\lambda \), which bears a 23 bp RSS (Nussenzweig, 1998; Nemazee and Weigert, 2000). Nevertheless, generation of functional IgH genes in mouse pre-B cell lines with nonfunctional IgH rearrangements can proceed by \( V_\lambda \) replacement (Kleinfield et al., 1986; Reth et al., 1986; Covey et al., 1990; Usuda et al., 1992). Comparison of the functional VDJ with the original nonfunctional VDJ rearrangement suggests that \( V_\lambda \) replacement may be mediated through the use of cryptic RSS (cRSS) sequences located within the third framework region of the \( V_\lambda \) germline gene segments (Kleinfield et al., 1986; Reth et al., 1986). The biological potential for \( V_\lambda \) replacement is supported by studies of knockin mice. Self-reactive VDJ genes artificially inserted into the \( J_\lambda \) locus can be altered by secondary rearrangements, including \( V_\lambda \) replacement (Chen et al., 1995a, 1995b, 1997). In humans, 40 out of 44 functional \( V_\lambda \) germline genes contain cRSS motifs with a heptamer (TACTGTG) but no clearly definable nonamer partner (Radic and Zouali, 1996). This cRSS motif can be found within the 3' end of \( V_\lambda \) genes of all vertebrates that have been examined, including mammals, birds, and cartilaginous fishes. Besides the cRSS within the third framework region of \( V_\lambda \) germline genes, other cRSS-like motifs containing CAC sequences may be involved in \( V_\lambda \) gene revision (Wilson et al., 2000; Itoh et al., 2000; Nemazee and Weigert, 2000). However, the molecular mechanism of cRSS usage in RAG-mediated recombination and the biological significance of \( V_\lambda \) replacement still need elucidation.

The EU12 cell line, established from a childhood acute lymphocytic leukemia patient (Zhou et al., 1995), undergoes continuous pro-B to pre-B and B cell differentiation (Wang et al., 2003). The RAG1, RAG2, and TdT genes are expressed at the pro-B stage, but TdT expression is extinguished at the pre-B cell stage, and RAG1 expression is attenuated as cells reach the immature B cell stage. Analysis of the BCR repertoire generated in this cell line has indicated intraclonal diversification of \( V_\kappa \) and \( V_\lambda \) genes (Wang et al., 2003). Here we describe studies demonstrating that this intraclonal IgH diversity is mediated by serial \( V_\lambda \) gene replacement. This dynamic model system allowed us to dissect the molecular mechanism of the ongoing \( V_\lambda \) gene replacement. The information obtained in this analysis also allowed us to examine the role of \( V_\lambda \) gene replacement in generating the normal B cell repertoire in humans.

Results

Intraclonal \( V_\lambda \) Gene Diversification through Serial \( V_\lambda \) Replacement in a Human B Cell Line

All of the IgH sequences expressed by the B cell members of the EU12 cell line were found to include the same \( D_{\lambda 0-10} J_{\lambda 0} \) joint but to employ different \( V_\lambda \) genes. The \( V_\lambda \) genes expressed by the EU12 cells, \( V_{\lambda 0-3}, V_{\lambda 0-6}, \) and \( V_{\lambda 0-11} \), are clustered within the D\( _\lambda \) proximal region of the heavy chain locus (Figures 1A and 1B). The prevalent IgH cDNA clones contain relatively short N region additions at their \( V_\lambda -D_\lambda \) junctions, while other clones contain longer N region additions (Wang et al., 2003). Inspection of the \( V_\lambda -D_\lambda \) junctions led to the identification of common motifs shared by the different sequences. All three groups of \( V_\lambda \) sequences contain CACA and TGGACT motifs upstream of the \( D_{\lambda 0} \) joint (Figure 1C), but these are associated with different \( V_\lambda \) genes and have variable CDR3 lengths. None of these motifs matched any of the \( D_\lambda \) gene sequences and their recurrence in multiple \( V_\lambda -D_\lambda \) joints made them unlikely to be N nucleotides. The first clue to their possible origin came from the identification of a nonfunctional \( V_{\lambda 0-5} D_{\lambda 0} J_{\lambda 0} \) rearrangement in the EU12 cells by single-cell PCR amplification of genomic DNA. This \( V_{\lambda 0-5} D_{\lambda 0} J_{\lambda 0} \) rearrangement had a \( D_{\lambda 0-10} J_{\lambda 0} \) joint identical to that observed in all of the functional IgH sequences. The 3' end of the \( V_{\lambda 0-5} \) germline gene contains the CACA motif, thereby suggesting this could be the origin of the CACA motif common to all three groups of sequences (Figure 1C). This realization led to the recognition that the CGAGAG motif in the \( V_{\lambda 0-1} D_\lambda \) rearrangement is identical to the sequence of the 3' end of the \( V_{\lambda 0-7} \) gene, while the CAAAAG motif found in the \( V_{\lambda 0-11} D_\lambda \) rearrangement is identical to the 3' end of the \( V_{\lambda 0-9} \) gene (Figure 1C). These findings suggested a model of EU12 intraclonal diversification through serial \( V_\lambda \) replacement mediated by the cRSS (TACTGTG) embedded within the 3' ends of the FR3 of \( V_\lambda \) genes (Figure 1D). Starting from a nonfunctional \( V_{\lambda 0-5} D_{\lambda 0} J_{\lambda 0} \) rearrangement, diverse functional and nonfunctional \( V_\lambda \) germline genes could be generated sequentially in progeny cells through serial \( V_\lambda \) replacement. With each round of replacement, the resulting IgH gene would obtain a new \( V_\lambda \) coding region from an upstream \( V_\lambda \) germline gene while retaining a short stretch of nucleotides from the 3' end of the \( V_\lambda \) gene used previously. In keeping with this hypothesis, more than half of the IgH sequences obtained from EU12 cells could be assigned as serial \( V_\lambda \) replacement products of different \( V_\lambda \) replacement rounds. This model also predicts that the dominant IgH cDNA clones with short N regions were likely generated during early steps of \( V_\lambda \) replacement, and the less abundant clones would have been gener-
Figure 1. Experimental Model for Serial $V_H$ Replacement in EU12 Cells

(A) Phenotypic analysis of EU12 cells defines three distinct subpopulations. The differentiation stages of the EU12 cells and the expression patterns for the TdT, RAG1, and RAG2 genes are illustrated. R3A cells express pre-BCR and R3B cells express BCR.

(B) Relative genomic location of $V_H$ genes segments rearranged in the EU12 cells.

(C) EU12 IgH sequences that may be derived from serial $V_H$ replacement. Color-coded sequences indicate theoretical origins for shared common motifs. The cRSS heptamer is boxed.

(D) Proposed model of serial $V_H$ replacement beginning with a nonfunctional $V_H_{2-5}J_H$ rearrangement to generate multiple functional IgH genes. Color-coded sequences indicate theoretical origins for shared common motifs. The boxes indicate the cRSS heptamers. The red arrowheads indicate the predicted cleavage sites.
RSS and 23 bp RSS also blocked protein/cRSS complex formation in a cross competition assay, suggesting that these DNA/protein complexes contain RAG proteins (see Supplemental Figure S1A at http://www.immunity.com/cgi/content/full/19/1/21/DC1). When the four probes derived from representative V_{H1-8}, V_{H2-5}, V_{H3-7}, and V_{H4-4} cRSS were tested in the EMSA study, similar protein/DNA complexes were observed (see Supplemental Figure S1B).

To confirm RAG protein binding to cRSS, we used purified recombinant GST-RAG1 and GST-RAG2 core proteins to perform in vitro binding assays. Binding to the V_{H1-8} cRSS probe was observed to depend upon the presence of both RAG-1 and RAG-2 (Figure 3A); neither protein alone was able to associate with the probe. Notably, efficient binding to the cRSS probe requires the coexpressed, copurified RAG-1/RAG-2 proteins, while the mixture of independently expressed purified RAG-1 and RAG-2 proteins has very low binding activity. This result is concordant with previous analyses of RAG protein binding to the 12 or 23 bp spacer RSS (van Gent et al., 1996; Eastman et al., 1996; Hiom and Gellert, 1997, 1998), suggesting that a similar folding status is required for RAG-1/RAG-2 binding to all types of RSS. The addition of a monoclonal anti-GST antibody specifically induced supershifting of the RAG-1/RAG-2/cRSS complexes as a further confirmation of the composition of these complexes (Figure 3B). To verify the generality of RAG binding to cRSS, we used cRSS probes derived from the V_{H1-8}, V_{H2-5}, V_{H3-7}, and V_{H4-4} genes. Recombinant RAG-1/RAG-2 was found to bind to all four cRSS, which are representatives of the cRSS present in most V_{H} germ-line genes (Figure 3C).

The binding affinity of recombinant RAG proteins to the V_{H1-8} cRSS probe was assessed by cross competition with increasing concentration of unlabeled V_{H1-8} cRSS, 12 bp RSS, 23 bp RSS probes, and a nonspecific control probe (Figure 3D). Based on this analysis, the binding affinity of the RAG proteins to the cRSS is approximately 5-fold less than that of RAG binding to the conventional 12 or 23 bp RSS.

Recombinant RAG-1/RAG-2 Proteins Cleave cRSS Sites in Vitro

The binding of RAG-1/RAG-2 proteins to the cRSS suggests that cRSS can participate in RAG-mediated recombination. To test this hypothesis, we performed in vitro cleavage assays using recombinant RAG-1 and RAG-2 proteins and cRSS probes derived from the V_{H1-8}, V_{H2-5}, V_{H3-7}, and V_{H4-4} genes as substrates (Figures 4A and 4B). Nicking was observed at the heptamer ends of all four cRSS probes with the addition of Mg^{2+}, but not with Ca^{2+}. The formation of coding end hairpin structures was also observed with all four probes when Mn^{2+} was included in the cleavage reaction. The cleavage patterns were virtually identical to those observed for conventional 12 or 23 bp RSS substrates under the same reaction conditions, except that RAGs cleave the cRSS sites less efficiently. Additional nicking was also observed for the V_{H1}, V_{H2}, and V_{H3} cRSS probes, and the size of the products predicts cleavage at the internal CAC sites. For the four cRSS, RAG-mediated nicking occurred at

---

**Figure 2. Ongoing V_{H} Replacement Recombination Events in EU12 Cells**

(A) Schematic illustration of predicted serial V_{H} replacements. The red arrowheads indicate the predicted cleavage sites.

(B) Detection of RAG-mediated double-stranded DNA breaks at the V_{H2-5} cRSS site using the ligation-mediated PCR technique. The + and − symbols indicate the presence or absence of T4 DNA ligase in the ligation reaction. The Link1 and LinkRSS primers that were used in the LM-PCR are indicated. Genomic DNA was prepared from either EU12 (lanes 1 and 2) or Jurkat cells (lanes 3 and 4).

(C–E) Nested primer PCR detection of the excision circles predicted for the V_{H2-5}→V_{H3-7} (C), V_{H3-7}→V_{H1-8} (D), and V_{H1-8}→V_{H3-9} (E) replacement events. Lane 1, EU12 cells; lane 2, Jurkat cells. All PCR products were sequenced with V_{H} specific primers. P indicates predicted sequences, and S13, C3, S4, and S6 indicate experimentally determined sequences.
Figure 3. RAG Proteins Bind to cRSS

(A) Sequence alignment of cRSS from the VH1-8, VH2-5, VH3-7, and VH4-4 genes. The box indicates the cRSS heptamer. The heptamer and nonamer consensus sequences are shown in gray.

(B and C) Purified RAG-1, RAG-2 proteins associate with the VH2-5 (B) or VH1-8, VH2-5, VH3-7, and VH4-4 cRSS (C) probes. The black arrowheads indicate the migration positions of the RAG-1/RAG-2/cRSS complex.

(D) Cross competition assays. The VH2-5 cRSS was radiolabeled and used as a probe. Unlabeled VH1-8 cRSS, 12 bp RSS, 23 bp RSS were used at 1/1000, 5/1000, 25/1000, or 50/1000 molar excess for competition. The nonspecific probe was used at 50× molar excess for competition.

levels comparable to that seen for a canonical 12 bp RSS. The coding end hairpin structure formation was found to be more efficient for VH1 and VH3 cRSS, which occurred at approximately 50% and 30%, respectively, of the level observed with a 12 bp RSS (see Supplemental Figure S2). These results indicate that the cRSS are universally functional in the RAG-mediated in vitro cleavage reaction.

Coupled Cleavage Reaction between cRSS and the 12 or 23 bp Spacer RSS

The specific recombination of VH→DH or DH→JH is governed by the 12/23 rule, which relies on the different lengths of the RSS spacer regions. The 12/23 restricted recombination specificity choice is reflected in a coupled cleavage reaction in vitro, wherein the addition of unlabeled 12 bp RSS probes can enhance the coding end hairpin structure formation of the labeled 23 bp RSS probes and vice versa (McBlane et al., 1995; van Gent et al., 1996; Eastman et al., 1996; Hiom and Gellert, 1997, 1998). To evaluate the cRSS recombination specificity choice, we conducted coupled cleavage reactions using radiolabeled 12 bp RSS or 23 bp RSS as substrates with the addition of unlabeled 12 bp RSS, 23 bp RSS, a nonspecific oligonucleotide, or the four different cRSS as secondary oligonucleotides. With the 12 bp RSS substrate, enhanced hairpin formation was observed by addition of the 23 bp RSS, but not the nonspecific control.

Addition of VH1, VH2, and VH3 cRSS probes also enhanced hairpin formation (Figure 4C). In a reciprocal experiment with radiolabeled 23 bp RSS substrate, the addition of 12 bp RSS, but not the control oligonucleotide, enhanced hairpin formation. Once again, the addition of VH1, VH2, VH3, and VH4 cRSS probes enhanced hairpin formation (Figure 4D). These results suggest a relaxation of 12/23 restricted recombination when the cRSS are used, thereby permitting recombination of cRSS with either 23 or 12 bp RSS.

Evidence of Ongoing VH Replacement in Human Bone Marrow B Cells

The analysis of ongoing VH replacement in the EU12 cell line suggested a strategy to determine whether VH replacement is a normal feature of B cell development. Ligation-mediated PCR was used to detect double-stranded DNA breaks at the cRSS sites of VH genes at different stages during B lineage differentiation in the bone marrow. In these experiments, the B lineage cells were separated into pro-B, pre-B, immature B, naive B, and memory B cell subpopulations; non-B lineage cells (CD19−) served as a negative control (Figure 5A). A strong LM-PCR signal was detected with a VH-specific probe in the immature B cell subpopulation but was not seen at other stages in B lineage differentiation or in non-B lineage cells (Figure 5B). Sequence analysis of the LM-PCR products confirmed the participation of...
Figure 4. RAG Proteins Cleave crSS In Vitro
(A) Diagram of the in vitro cleavage assay indicating the experimental substrates and the cleavage products. The star indicates the [32P]-dCTP labeling position.
(B) Recombinant RAG-1/2 mediated cleavage of the 12 bp RSS and the crSS from the V\(_{\text{H}1-8}\), V\(_{\text{H}2-5}\), V\(_{\text{H}3-7}\), and V\(_{\text{H}4-4}\) genes in the presence of different divalent cations. U indicates uncut substrate, N indicates nicking product, and H indicates hairpin structure.
(C and D) Specific recombination of crSS with 12 bp or 23 bp RSS. The coupled cleavage reaction was performed using the 12 bp RSS (C) or 23 bp RSS (D) as substrates, with the addition of 12 bp RSS, 23 bp RSS, non-specific control probe, or the four crSS derived from V\(_{\text{H}1-8}\), V\(_{\text{H}2-5}\), V\(_{\text{H}3-7}\), or V\(_{\text{H}4-4}\) gene.

V\(_{\text{H}3-43}\) and V\(_{\text{H}3-66}\) genes in this step of the V\(_{\text{H}}\) replacement process (Figure 5C). Using V\(_{\text{H}1}\) or V\(_{\text{H}4}\) family-specific primers, LM-PCR products were also detected in the immature B cell population of bone marrow samples from other individuals (data not shown). This analysis suggests the occurrence of V\(_{\text{H}}\) replacement at the immature B cell stage in normal human bone marrow.

Evidence of V\(_{\text{H}}\) Replacement in Natural IgH Sequences
The demonstration that serial V\(_{\text{H}}\) replacement diversified the repertoire in the EU12 cells together with the evidence of V\(_{\text{H}}\) replacement in immature B cells in the bone marrow suggest that V\(_{\text{H}}\) replacement may contribute to the diversification of the primary B cell repertoire. In a search for V\(_{\text{H}}\) replacement products, we analyzed human IgH gene V-D joint sequences for "footprints" of formerly used V\(_{\text{H}}\) genes in the form of pentameric nucleotides that match the 3' end of germline V\(_{\text{H}}\) genes. Among the 343 functional IgH sequences (Zemlin et al., 2001) that were analyzed, 16 (~5%) contained at least one pentameric footprint at the V-D junction region that matched another V\(_{\text{H}}\) germline gene and thus could be designated as a potential V\(_{\text{H}}\) replacement product (Figures 6A and 6B). One of the potential V\(_{\text{H}}\) replacement products, C18a16, contains two such pentamers and could have been generated by two rounds of serial V\(_{\text{H}}\) replacement.

Figure 5. Ligation-Mediated PCR (LM-PCR) Analysis of Double-Stranded DNA Breaks at the crSS Sites in Bone Marrow Subpopulations of B Lineage Cells
(A) Purification of control and B cell subpopulations by FACS.
(B) Detection of double-stranded DNA breaks at the VH3 crSS site. The DNA blot of second round V\(_{\text{H}3}\) crSS LM-PCR products is hybridized with a V\(_{\text{H}3}\) crSS-specific probe. Lane numbers correspond to the gated cell fractions indicated in (A). The CD19 promoter region was amplified by PCR as a control for DNA input.
(C) Sequence analysis of the LM-PCR products confirms the occurrence of double-stranded DNA breaks at the V\(_{\text{H}3}\) crSS sites. The boxes indicate crSS heptamers. Linker primer sequences are underlined.
Using the same analysis stringency to examine the D-J junction region sequences, which would not be expected to contain VH replacement footprints, only two sequences were found to contain a pentamer that matched a VH germline gene. The significantly higher frequency of perfectly matched VH pentamers found in the V-D junctions versus D-J junctions (p = 0.009) supports the validity of this VH replacement search strategy. Two types of VH replacement events were suggested by analysis of potential VH replacement products. Among the 16 potential VH replacement products, 41% could have been generated through converting previously nonfunctional rearrangements into functional ones, while 59% were more likely generated by replacing one functional VDJ joint with another one.

In this analysis, the donors for VH replacement can be identified with certainty, since the entire VH gene segment is present in the resulting IgH sequence. The recipients of VH replacement contain sequence remnants that may be found in multiple VH germline genes. Therefore, the putative recipients were assigned as the closest 3' germline VH gene segment which could provide the footprint of VH replacement. The candidate VH replacement donors and recipients, plotted on the human VH gene locus map (Figure 6B), include VH germline genes spanning the locus. Consistent with the fact that 40 of the 44 human germline VH genes contain cRSS sites (Radic and Zouali, 1996), our data thus suggest that virtually the entire spectrum of cRSS could participate in RAG-mediated VH replacement. This analysis also suggests that VH germline genes that do not contain a cRSS, such as VH3-15, may serve as donors for VH replacement. Of the 16 potential VH replacement products, 12 could be generated using upstream VH genes to replace rearranged downstream VH genes, while 4 appear not to follow this order. The latter VH replacements could reflect inversion or transrecombination events, but more definitive interpretation of these results will require better resolution of the polymorphic complexity of the human VH locus.

The Footprints of VH Replacement Contribute Highly Charged Amino Acids to the IgH CDR3 Region

During each round of VH replacement in the EU12 cell line, the entire VH coding region is replaced by a new germline VH gene, except for a short stretch of 3' nucleotides that is left behind as a footprint within the newly formed IgH CDR3 region. Analysis of the amino acid sequences of the VH replacement products in the leukemic EU12 cells and normal B cells in humans yielded...
the remarkable finding that 18 of the 19 VH replacement footprints encode highly charged amino acids (Figure 7A). An analysis of the amino acids sequences at the 3’ ends of all VH germline genes that may be used as VH replacement footprint donors indicates that 80% of the potential codons would yield highly charged amino acids (Figure 7B). In striking contrast, only 10% of the amino acids encoded by germline DH genes are charged amino acids (Figure 7C). When the frequency of arginine, aspartic acid, and glutamic acid in the V-D junction of VH replacement products, the V-D junction of non-VH replacement IgH genes, and the D-J joints were compared, the frequency of these amino acids is significantly higher in the VH replacement products (Figure 7D). The addition of charged amino acids to the IgH CDR3 region in humans. It is noteworthy that the strategy used to search for residual footprints within the V-D junction thus appears to be a prominent feature of the VH gene replacement process.

Discussion

The concept of VH replacement was originally suggested by analysis of IgH genes expressed by murine pre-B cell lines in which some of the cells spontaneously converted from a μ heavy chain negative to a μ positive phenotype (Kleinfield et al., 1986; Reth et al., 1986). The identification of a cryptic RSS heptamer within the third framework region of most germline VH genes suggested that this could potentially serve as a recombinase target for the VH replacement reaction (Kleinfield et al., 1986; Reth et al., 1986; Covey et al., 1990). The present studies of the VH replacement mechanism and assessment of its contribution to the preimmune repertoire began with the analysis of IgH sequences derived from the EU12 human B lineage cell line, which undergoes in vitro differentiation from the pro-B to B cell stage (Wang et al., 2003). The repetitive occurrence of 6 nucleotide stretches within the V-D junction of IgH sequences from EU12 cells provided the first hint of a serial VH replacement process. Following this lead, direct evidence for ongoing RAG-mediated serial VH replacement was obtained initially in the clonal EU12 cells and then in primary B lineage cells. Moreover, evaluation of a large IgH sequence database indicated that VH replacement contributes significantly to the primary B cell repertoire in humans. It is noteworthy that the strategy used to search for residual footprints within the V-D junction could be applied to human IgH sequences because most human VH germline genes contain 6–7 nucleotides downstream of the conserved cRSS. This strategy is unfeasible in mice, wherein most VH germline genes contain fewer than 5 nucleotides after the cRSS.

VH Replacement Occurs through cRSS-Directed Recombination

Detection of double-stranded DNA breaks at the heptamer and identification of the expected excision circles suggested that VH replacement occurs through RAG-mediated recombination. In vitro studies using purified recombinant RAG-1/RAG-2 proteins provided evidence that RAG complexes are able to bind and cleave the
cRSS sites found in most V<sub>H</sub> genes. The reasonably efficient binding and cleavage of the cRSS by RAG complexes offer a molecular explanation for participation of the cRSS in RAG-mediated recombination. The in vitro relaxation of the specificity of RAG-mediated recombination is a well-recognized phenomenon, since a single heptamer motif or even a highly degenerative CAC motif can be cleaved by RAG (Ramsden et al., 1996; Kim and Oettinger, 1998; Swanson and Desiderio, 1999; Ragha-van et al., 2001). Our composite results extend these in vitro observations to provide a molecular explanation for efficient V<sub>H</sub> replacement mediated by cRSS containing a heptamer but no clearly identifiable nonamer. The specificity of primary V(D)J recombination is safeguarded by the 12/23 rule, as RSS with a 12 bp spacer can efficiently recombine only with a 23 bp spacer RSS and vice versa (Tonegawa, 1983; Lewis, 1994). This 12/23 restricted recombination substrate nucleotide setting is recapitulated in a coupled in vitro cleavage assay, in which the addition of a matched secondary probe (12 bp RSS for 23 bp RSS, or vice versa) enhances the coding end hairpin structure formation during the RAG-mediated cleavage (Eastman et al., 1996; van Gent et al., 1996; Hiom and Gellert, 1998). Interestingly, when cRSS probes are used in the in vitro coupled cleavage analysis, they enhance the hairpin structure formation of either 12 bp RSS or 23 bp RSS substrates. The cRSS have no definable nonamer and thus no specified spacer length, which may exempt them from the 12/23 restriction.

**Catching B Cells in the Act of V<sub>H</sub> Replacement**

Detection of double-stranded DNA breaks at the cRSS sites of the V<sub>H</sub> germline genes in the leukemic EU12 cells and normal bone marrow B cells provides direct evidence for the occurrence of V<sub>H</sub> replacement. The enrichment of cRSS double-stranded DNA breaks at the immature B cell stage is also consonant with previous studies in mice that show inducible expression of RAG genes and editing of the immunoglobulin light chain genes at the immature B cell stage (Casellas et al., 2001; King and Monroe, 2001). The evidence for both V<sub>H</sub> replacement and secondary light chain gene rearrangements at the immature B cell stage therefore suggests that this is the differentiation stage at which editing of antigen receptors occurs in humans. In mice carrying knockin self-reactive IgH VDJ regions, V<sub>H</sub> replacement is employed to delete a highly self-reactive IgH gene, presumably as a consequence of self-antigen encounter (Chen et al., 1995b). V<sub>H</sub> replacement could also occur in pro-B cells carrying nonfunctional IgH genes. This type of replacement is the initial replacement event in the EU12 cells and also is thought to occur in murine progenitor B cell lines (Kleinfield et al., 1986). Theoretically, V<sub>H</sub> replacement could also occur in pro-B cells producing IgH chains that do not pair well with the surrogate light chain. However, under the conditions employed in our studies, LM-PCR signals from double-stranded DNA breaks were not detected at cRSS sites in primary pro-B or pre-B cells, perhaps because of the infrequency of V<sub>H</sub> replacement at these early differentiation stages or because the double-stranded DNA breaks are quickly repaired in cells undergoing extensive proliferation.

**Contribution of V<sub>H</sub> Replacement to the Human B Cell Repertoire**

Beginning with a nonfunctional V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> joint, continuous serial V<sub>H</sub> replacement diversifies the V<sub>H</sub> repertoire in the EU12 cells. The serial V<sub>H</sub> replacement not only rescues pro-B cells carrying a nonfunctional V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> but also generates intraclonal diversity. Importantly, with each round of V<sub>H</sub> replacement, the resulting IgH gene renews the entire V<sub>H</sub> coding region except for a short stretch of 3’ nucleotides from the replaced V<sub>H</sub> gene that is retained in the V<sub>i</sub>-D<sub>i</sub>-J<sub>i</sub> join. This residual sequence serves as a diagnostic marker that can be used to search for potential V<sub>H</sub> replacement products in primary B cells. Through an analysis of IgH gene sequences derived from normal individuals of different ages, we could identify potential V<sub>H</sub> replacement products in approximately 5% of analyzed sequences using a stringency setting of 5 nucleotide matches with no mismatch. With a less stringent setting, allowing 1 mismatch among 6 conserved nucleotides, we could identify potential V<sub>H</sub> replacement products in approximately 12% of analyzed sequences (data not shown). While this figure represents a very significant contribution to the repertoire, the actual frequency of V<sub>H</sub> replacement is likely to be even higher, given that a third of the functional V<sub>H</sub> gene alleles have less than 6 nucleotides downstream of the cRSS motif. The V<sub>H</sub> 3’ end sequences that we depended upon for assignment of the potential V<sub>H</sub> replacement product may also be nibbled on either end by exonuclease activity during coding joint formation or mutated by somatic hypermutation. Using the perfect pentamer criterion, we identified a single instance in which the sequence may reflect two rounds of V<sub>H</sub> replacement. This suggests that multiple rounds of V<sub>H</sub> replacement may either be rare in primary B cells or their footprints are obscured through subsequent changes in the V<sub>H</sub> region. With regard to the frequency of V<sub>H</sub> replacement, with even the most conservative interpretation of the data we estimate at least 1 in 20 B cells undergoes V<sub>H</sub> replacement.

**Biological Consequences of V<sub>H</sub> Gene Replacement**

The ability to replace V<sub>H</sub> genes that are dysfunctional or highly self-reactive would be the most obvious biological advantage of the V<sub>H</sub> replacement process. The paradoxical finding that V<sub>H</sub> replacement preferentially adds charged amino acids to the CDR3 region was therefore surprising. Yet, this is a consistent feature of V<sub>H</sub> gene replacement in that the 3’ ends of all human and mouse V<sub>H</sub> germline genes encode primarily charged amino acids irrespective of their reading frame. This appears to be a unique contribution to the IgH CDR3 region since charged amino acids are rarely encoded by human or mouse germline D<sub>H</sub> genes or by N region sequences at the D-J joints. The biological significance of adding these charged amino acids to the CDR3 region is presently uncertain. Charged amino acids within the CDR3 region has been identified as a hallmark of autoreactive antibodies. In particular, arginine is found frequently within the IgH CDR3 regions of anti-DNA antibodies (Radic et al., 1993) and almost one-third of the potential codons derived from the footprints of V<sub>H</sub> replacements contribute an arginine residue. Whether V<sub>H</sub> replacement contributes to the generation of self-reactive antibodies
remains to be determined. An alternative possibility is that these charged amino acids in the CDR3 region contribute to low-affinity self-reactivity that would promote clonal survival rather than harmful consequences.

**Experimental Procedures**

**Cell Culture**

EU12 cells grown under standard tissue culture conditions were routinely monitored for surface CD34 and μHC expression by FACS (Wang et al., 2003).

**Ligation-Mediated PCR**

Ligation-mediated PCR (LM-PCR) employed a modified procedure used to detect double-stranded DNA breaks at conventional RSS sites (Schlissel et al., 1993). In brief, genomic DNA (1 μg) was ligated to 20 pmol of reannealed double-stranded linker (Schlissel et al., 1993), and double-stranded DNA breaks were detected with two rounds of nested PCR. For the first round, 10 μl of the ligation reaction was used as template and amplified with sense primer Vμ together with an antisense primer Linker1. For the second round PCR, 2 μl of the first round PCR reactions was used as template and amplified with an internal Vμ, sense primer and the LinkRSS antisense primer. Second round PCR products (10 μl) were separated on 1% agarose gels and visualized with EtBr staining. Gel-purified PCR products (QiaExII, Qiagen) were directly sequenced using the Vμ internal primer. For detection of double-stranded DNA breaks at the VμRSS from human bone marrow cells, the conditions were similar except that the LinkRSS primer was used in the first round PCR and LinkRSS or LinkcRSS was used in the second round PCR to increase the reaction stringency. Second round LM-PCR products were purified and subcloned into the pCRII vector, and positive clones were selected for sequencing to verify the location of double-stranded DNA breaks. Ligation reaction samples were subjected to PCR amplification of the CD19 promoter region to monitor the DNA input.

**Detection of Vμ Replacement Excision Circles**

The method was adopted from procedures originally designed for amplifying IgD recombination excision circles (Hikida and Ohmori, 1998). In brief, cellular DNA was prepared from 10⁶ EU12 cells and subjected to two rounds of PCR amplification with seminested primer sets. The second round PCR products were separated by 2% agarose gel electrophoresis and visualized by EtBr staining and sequenced as above.

**Electrophoretic Mobility Shift and In Vitro Cleavage Assays**

EMSA was performed as described (Hiom and Gellert, 1997). Nuclear extracts were prepared from EU12 cells. The cRSS probes were derived from Vμ-RSS, Vμ-RSS, Vμ-RSS, and Vμ-RSS genes and labeled with [³²P]-dCTP. For competition assays, the unlabeled cRSS, conventional 12 bp RSS, 23 bp RSS, or nonspecific probe were used at 50-fold molar excess relative to the radiolabeled probe. HMG-1, GST-fusion RAG-1, and RAG-2 core proteins were purified as described elsewhere (Eastman et al., 1999; Hiom and Gellert, 1998; Huyse and Roth, 2000). For in vitro binding assays, purified RAG-1 and RAG-2 proteins were used at 10 ng per binding reaction in EMSA binding buffer as described (Hiom and Gellert, 1997). For supershift assays, 0.2 μg of monoclonal anti-GST antibody (Sigma) or mouse IgG was added into the binding reaction prior to the addition of radiolabeled probes and incubated for 30 min at 4°C. The binding reactions were resolved on 5% native PAGE gels with 1× TBE at 4°C. Dried gels were exposed to X-ray films at ~70°C for different time intervals.

The in vitro cleavage was performed in cleavage buffer (Hiom and Gellert, 1997) containing 20% DMSO and supplemented with Ca²⁺ (5 mM), Mg²⁺ (5 mM), or Mn²⁺ (1 mM). For coupled cleavage, the secondary oligonucleotides were used at 5-fold molar excess for 12 bp RSS or 23 bp RSS and 20-fold molar excess for all the cRSS. Cleavage was performed as follows: after 10 min incubation of radiolabeled probes (20 fmol), RAG1 (100 ng), RAG2 (100 ng), HMG-1 (20 ng), and secondary oligonucleotides in binding reaction buffer containing 5 mM of Ca²⁺ at 37°C, cleavage was initiated with the addition of 5 mM Mg²⁺ and continued for 90 min at 37°C. The cleavage reaction was stopped by the addition of an equal volume of loading buffer containing 90% formamide and 10 mM EDTA. Samples were heated for 10 min at 95°C and separated at 4°C on 20% sequencing gels containing 8 M urea. Gels were briefly fixed in 30% acetic acid/10% methanol, rinsed with distilled H₂O, dried, and exposed to X-ray film at ~70°C.

**Purification of Bone Marrow B Cell Subsets**

Human bone marrow mononuclear cells were prepared as described elsewhere (Nunez et al., 1996). The B cell subsets were purified by FACS using the following phenotypic markers: pro B cells (CD19⁺ CD34⁻ μHC⁻), pre B cells (CD19⁺ CD34⁻ μHC⁻), immature B cells (μHC⁺ iHC⁺ CD24⁻), naive B cells (μHC⁺ iHC⁻ CD24⁺), and memory B cells (μHC⁺ CD24⁺) (Nunez et al., 1996).

**Sequence Analysis of IgH Genes**

From 413 Igh sequences with unique CDR3 (Zemlin et al., 2001) (GenBank accession numbers AF235505–235921), 343 sequences containing an unambiguously identifiable Dμ gene segment were analyzed to identify potential Vμ replacement products. In brief, the IgV-D-J joint regions were analyzed for tetramer motifs, which are identical to the 3' end of other Vμ germ-line genes (see Supplemental Data at http://www.immunity.cgl.org/content/full/1/1/21/DC1). For the potential Vμ replacement product, the existing Vμ germ-line gene was assigned as the Vμ replacement donor. The Vμ replacement recipients were predicted based on the possible origin of the pentamers found in the V-D-J joint regions. The identification of the functional or nonfunctional rearrangement status of the previous VDJ joints was based upon the reading frame of the identified pentamers within the V-D-J joint regions.

**Acknowledgments**

We would like to thank Drs. Robert Stephan, Harry Schroeder, Chen-lo Chen, and Hiromi Kubagawa for helpful discussions, and Dr. Larry Gartland for FACS sorting. This work was supported in part by NIH grants AI39816 and AI48098. Z.Z. is supported in part by a K01 grant AR48592 from NIH/NIAMS. M.Z. is supported by a Humboldt-Foundation grant FLF1071857. D.B.R. is an Assistant Investigator and M.D.C. is an Investigator of The Howard Hughes Medical Institute.

Published: July 15, 2003

**References**


