Preferential Utilization of Specific Immunoglobulin Heavy Chain Diversity and Joining Segments in Adult Human Peripheral Blood B Lymphocytes

By Masao Yamada,* Robert Wasserman,† Betty Anne Reichard,* Sara Shane,* Andrew J. Caton,* and Giovanni Rovera*

From *The Wistar Institute of Anatomy and Biology; and The †Department of Pediatrics, The University of Pennsylvania School of Medicine and The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

Summary

We have examined at the molecular level the CDR3 and adjacent regions in peripheral blood B lymphocytes of normal individuals. A total of 111 sequences (12–28 sequences from six individuals) were obtained after cloning of the polymerase chain reaction–amplified segments into plasmids or phage. The average length of the VDJ joining was 109 nucleotides, with a range from 79 to 151. Approximately 75% of the sequences were in frame when translated into amino acids. Among the JH segments, JH4 was found most frequently (in 52.5% of the sequences), and JH1 and JH2 segments the least frequently (~1% of the clones). A polymorphic JH6 gene with a one-codon deletion accompanied by a base change was present in two of six patients. Preferential breakpoints were found for JH2, JH3, JH4, and JH5, although the breakpoints of JH6 were distributed more heterogeneously.

In ~90% of the cases, significant homology of the D regions with published D sequences was found. Preferential usage of a particular coding frame was observed in in-frame sequences utilizing DA, D21/9, and DM1 segments. However, in general, all coding frames of germline D genes were used to generate CDR3s. Eight sequences that have a DN1-like D sequence with two base changes at the same positions were identified, suggesting the likely existence of a new germ line D gene belonging to the DNfamilies. Using probes specific for a particular CDR3, the frequency of a specific B cell clone in the peripheral blood of normal individuals was estimated to be at most as high as 1/20,000.

The most variable region of the immunoglobulin heavy chain is the third complementarity determining region (CDR3) (1, 2). This region spans the junction between the variable (V<sub>H</sub>) diversity (D), and junctional (J<sub>H</sub>) segments in the rearranged IgH genes (1, 2). The hypervariability of this region is due to the combinatorial assortment of the many V<sub>H</sub>, D, and J<sub>H</sub> segments that are utilized to generate a particular CDR3, to the imprecise joining mechanisms that include deletion of bases from the potential coding regions of each segment to be joined (3), and the addition of new bases that can be enzymatically added at the point of joining (N regions) (4). Finally, somatic mutations of the rearranged region can contribute to the production of higher affinity antibodies (5).

The nucleotide sequences of all the human D genes, estimated to be ~30 in number (6), have not been fully defined, and questions remain about the relative usage of different D genes used in VDJ joinings during development and in adult individuals. Furthermore, the characteristics of the extent of base excision and addition, including the identification of preferred sequence boundaries for the V<sub>H</sub>, D, and J<sub>H</sub> regions, have not been well delineated. The relative frequency of in-frame translation products reflecting productive rearrangements, and the possibility of specific translation frames being preferred for particular D gene families have not been determined on a large sample size.

To address these questions, we have used the method of the PCR (7, 8) using primers for framework region (3) (FR3) of the V<sub>H</sub> segments and for the J<sub>H</sub> genes to amplify the CDR3s and adjacent regions (9) from a polyclonal population of peripheral blood B lymphocytes. The amplified CDR3s have been sequenced from plasmids or phage libraries.

The resulting analysis of 111 CDR3 sequences from six adult volunteers gives a picture not only of the usage of various D family and J<sub>H</sub> gene segments and of the translational frame used in individual D family genes, but also extends...
our present information on the location of joining boundaries, the frequency of in-frame products in circulating B cells of normal individuals, and the type of polymorphisms present in the human JH gene locus. Also, we provide further evidence that novel rearrangements (Vr-Jx, Vr-inverted D-Jx, Vr-DIR-Jx, Vx-D-D-Jx, Vx-D-inverted D-Jx, Vx-DIR-D-Jx, Vx-inverted DIR-D-Jx, and Vx-DIR-Jx) are commonly present in the circulating B cell population supporting the original hypothesis of Kurosawa and Tonegawa (3), that alternative signal sequences present in certain D segments may be responsible for this break in the 12/23-bp spacer recombination rule (10).

**Materials and Methods**

**Clinical Samples and DNA Preparation.** Peripheral blood (10 ml) was obtained by venipuncture from six healthy adult volunteers. Low-density mononuclear cells were obtained by fractionation on a Ficoll/Hypaque gradient (d=1.078) (11). High molecular weight genomic DNA was isolated from mononuclear cells using established methods (12).

**Oligonucleotide Primers and Probes.** Oligonucleotides were synthesized by the solid phase triester methodology on a DNA synthesizer (380A; Applied Biosystems, Inc., Foster City, CA) (13). The sense and antisense primers containing the Sall and PstI cloning sites, homologous to the Vx, FR3 and the Jx genes, used for amplification of the VDJ joining region of the IgH, were 5'-CTG-TCCACACAGGCGGCTTATACTG-3' and 5'-AACCTGCAAGAGAGACGGTGACC-3', respectively. In some experiments, to exclude a possible bias in amplification of the Jx segments, primers specific for J2 and J3 were used. These J2 and J3 primers differed from the antisense universal Jx primer by a G to A base substitution at positions 16 and 9 of the universal Jx primer, respectively. The Jx consensus (JxC) probes, which were used to detect all VDJ rearrangements, were a mixture of four oligonucleotides derived from the sequences of J2, J3, Jx4, and Jx6 genes just 5' to the Jx antisense primer used for PCR amplification. These oligonucleotides had the following sequences; J2: 5'-CTG-GGCCCCTGGAACCTGGC-3'; J3: 5'-CTGAGGAAGGGA-CAATG-3'; Jx4: 5'-CTGAGGAAGGGAACCTGGC-3'; Jx6: 5'-TCTGGGAGGGAACCTGGC-3'.

The J4 probe also hybridized to Jx1 and Jx5 sequences when 42°C and 45°C were used as the hybridization and washing temperatures, respectively.

**PCR, Cloning, and Sequencing of VDJ Joining.** PCR was carried out as described by Saiki et al. (7) and Mullis and Faloona (8). The initial denaturation step was at 95°C for 5 min, followed by 30 cycles with a 1-min annealing step at 55°C, a 2-min elongation step at 70°C, and a 2-min denaturation step at 95°C. The final cycle was completed with a 7-min elongation step. Samples were extracted with phenol/chloroform, precipitated with ethanol, and resuspended in Tris-EDTA buffer. Precautions against cross-contamination of amplified material were taken according to the recommendations of Kwock et al. (14).

After phenol/chloroform extraction and ethanol precipitation, an aliquot of the PCR-amplified material was digested with both Sall and PstI restriction endonucleases (Boehringer Mannheim Biochemicals, Indianapolis, IN). After electrophoresis in a 4% NuSieve agarose gel (FMC, Rockland, ME), slices spanning the 72-191-bp region, which contain the amplified CDR3 DNA, were treated with agarase (15) (Calbiochem-Behring Corp., San Diego, CA). Recovered DNA was ligated into Bluescript phagemid (Stratagene, La Jolla, CA) and transfected into Escherichia coli strain JM 109 (16).

Transformants were lifted onto nitrocellulose filters (BA 85; Schleicher & Schuell, Inc., Keene, NH). The filters were hybridized at 42°C with 32P-labeled JxC probes and were washed in 6× SSC and 0.1% SDS at 45°C. Lower stringency of washing was used in order not to miss any VDJ clones with polymorphisms in the areas homologous to the JxC probes used.

Positive clones were picked up randomly, and double-stranded DNA template was prepared and sequenced by the method of Sanger et al. (17) using the Sequenase kit (United States Biochemical Corp., Cleveland, OH).

**Computer Analysis of DNA Sequences.** Computer analysis of DNA sequencing data was performed using the sequence analysis software package of the Genetic Computer Corp., Release 5, of the University of Wisconsin and a Micro Vax II computer (Digital Equipment Corp., Marlboro, MA).

All VDJ joining sequences were entered using the "sequed" program. Each sequence was checked for the presence, in the correct orientation, of the Jx and Vx primers. Each VDJ joining sequence was translated into the predicted amino acids, using the following criteria to identify in-frame sequences. (a) The amino acid sequence of the Vx primer in FR3 starts with VTVWYY (C). (b) The amino acid sequence of the Jx primer ends up with VTVSS (A). (c) No stop codons should be present in between the two primer coded sequences. Each sequence was searched for homology with the six human Jx sequences (18) available in a subdirectory using the "word search" program.

Each sequence was searched for homology against all published D genes (6, 19, 20) available in a second subdirectory. The best-fit D genes were chosen according to the following criteria; (for Fig. 7, we dropped criteria a and e). (a) We did not accept reverse homology. (b) We gave priority to the homology with a Jx segment, when both D and Jx homologous regions overlapped. (c) When we had more than one candidate, we took the longest stretch with the highest homology. (d) We tried to avoid introducing gaps and base additions to the alignment. However, if there were no other candidates and the stretch of homology was long enough, we took the sequence with the gap or base additions. (e) We did not use homology to DIR sequences. According to this computer analysis, sequences of the VDJ joinings examined were subdivided into Vx, D, and Jx segments.

**Generation and Screening of M13 Libraries of VDJ Joining Clones.** PCR-amplified materials from donors 2, 3, and 4 were ligated into M13 mp19 (Bethesda Research Laboratories, Gaithersburg, MD) phage vector using the same protocol used for Bluescript transformation. Ligated materials were then used to transform DH5α E. coli, and ~5-10 × 105 plaques per 15-cm plates were transferred to nitrocellulose filters to generate replica filters. Duplicate filters were screened separately with the JxC probes to establish the number of VDJ joinings present in the libraries and with diagnostic oligonucleotide probes homologous to the N regions of randomly picked VDJ joinings to establish the frequency of a specific CDR3 sequence in the overall population.

The frequency of J2 segment usage in the population was determined by screening separately a large number of M13 clones containing VDJ joinings on duplicate filters with the JxC probes that hybridize to all six Jx genes and with the J2-specific probe. Stringent hybridization conditions (washing the filters at 1°C below the melting temperature of the probe) were used with the J2 probe in order to avoid crosshybridization with other Jx segments. 12 of the plaques that hybridized to the J2 probe were isolated and sequenced.
Results

Characteristics of the CDR3 in Normal B Lymphocytes. DNA sequences that contain the CDR3 and adjacent regions spanning from the FR3 of V\textsubscript{\alpha} to the 3' end of J\textsubscript{\alpha} were obtained from PBL of six healthy adult volunteers.

Fig. 1 contains the sequences of 99 randomly picked VDJ joinings as well as 12 VDJ joinings containing J\textsubscript{2} segments that have been isolated using a J\textsubscript{2}-specific diagnostic probe. Palindromic nucleotides generated during the joining process, as described by Lafaille et al. (21), are shown when present as full tetrameric palindromes. In the 111 total CDR3 sequences, nine were found at the 3' V\textsubscript{\alpha} border, five at the 5' D border, one at the 3' D border, and one at the 5' V\textsubscript{\alpha} border.

Fig. 2 shows the predicted amino acid sequences of the CDR3 and adjacent regions. 75% (75/99) of the randomly picked sequences were in-frame when translated into amino acids. 7 of 12 VDJ joinings with J\textsubscript{2} segments (58%) were in-frame. The in-frame sequences from Fig. 1 are grouped according to the J\textsubscript{\alpha} utilized and are listed in an increasing order of length. The reading frame for each D segment is also indicated.

Fig. 3 shows the length distribution of the 99 randomly picked VDJ joinings shown in Fig. 1. Clone 3-79 (Fig. 1) is the shortest CDR3 identified, and it is an in-frame sequence that codes for only a four-amino acid-long CDR3 (Fig. 2). Clone 1-139 (Fig. 1) is the longest in-frame sequence that codes for a 24-amino acid-long CDR3 (Fig. 2). Clone 1-151 (Fig. 1) is the longest out-of-frame sequence. Although the length of the CDR3s are very heterogenous, ranging from 4 to 24 amino acids, the majority of clones range from 8 to 18 amino acids. Only one of four clones that were longer than 136 bp are in-frame, in contrast to the smaller sequences in which the majority are in-frame.

Characteristics of the J\textsubscript{\alpha} Regions Utilized in Adult B Lymphocytes. Table 1 shows the frequency of J\textsubscript{\alpha} gene usage in adult peripheral lymphocytes calculated from the 99 VDJ sequences shown in Fig. 1, which were amplified with the J universal primer. The most frequently found J\textsubscript{\alpha} segments are J\textsubscript{3} (52.5%), J\textsubscript{6} (22.2%), and J\textsubscript{5} (15.2%). Only one VDJ clone carrying a J\textsubscript{1} segment and no clones containing a J\textsubscript{2} segment were identified among the 99 clones randomly picked.

Experiments were done to rule out the possibility that the J universal primer was biased in its amplification of J\textsubscript{2} and J\textsubscript{3} segments due to one-base mismatch with these sequences. Three J primers 100% homologous to J\textsubscript{2}, J\textsubscript{3}, and J\textsubscript{4} were synthesized and used individually to amplify the CDR3 of the PBL of donor 3. The amplified products resulting from each J primer were cloned in Bluescripts and screened with probes specific for J\textsubscript{2}, J\textsubscript{3}, and J\textsubscript{4}. The data (not shown) indicate that no differences in the frequency of J\textsubscript{2}, J\textsubscript{3}, and J\textsubscript{4} containing recombinant clones were observed when either the J universal or J\textsubscript{3} or J\textsubscript{2} primers were used in the PCR.

Table 2 shows the frequency of J\textsubscript{2} gene usage as determined by screening an M13 library containing CDR3-amplified sequences with a J\textsubscript{2}-specific oligonucleotide probe (see Materials and Methods section). The calculated frequency of J\textsubscript{2} gene usage ranged from 0.98 to 1.77% in the three donors studied.

Fig. 4 shows the sequences present in the six donors in J\textsubscript{3}, J\textsubscript{4}, J\textsubscript{5}, and J\textsubscript{6} gene segments as compared with the sequences originally published by Ravetch et al. (18). These variant sequences have all been published previously (18, 22, 23). J\textsubscript{3}b differs from J\textsubscript{3}a by one base (G to A), which results in an amino acid change from V to I. J\textsubscript{5}b differs from J\textsubscript{5}a by three bases (C to A, A to G, T to C), of which the latter base change results in an amino acid change from S to P. J\textsubscript{6}c has three consecutive base deletions (GGT), which eliminates an amino acid, and one base change (C to A), which results in an amino acid change from Q to K when compared with J\textsubscript{6}b. J\textsubscript{4}b and J\textsubscript{6}b differ from J\textsubscript{4}a and J\textsubscript{6}a by one base change, respectively, which does not result in an amino acid change.

J\textsubscript{3}b, J\textsubscript{4}b, and J\textsubscript{5}b segments were exclusively used by all six donors. For J\textsubscript{3} and J\textsubscript{5}, there were one and three sequences, respectively, which could not be assigned to an a or b sequence because the break point was downstream from where they differed. These findings and other findings in the literature (23) indicate that J\textsubscript{3}b, J\textsubscript{4}b, and J\textsubscript{5}b sequences should be considered as prototype sequences instead of J\textsubscript{3}a, J\textsubscript{4}a, and J\textsubscript{5}a sequences, which were originally reported as germline J\textsubscript{\alpha} sequences in a single individual (18). All the J\textsubscript{6} sequences obtained from donor 6 used the J\textsubscript{6}c sequence. This variant sequence was also observed in two of three J\textsubscript{6} sequences from donor 5, while J\textsubscript{6}b sequences were found exclusively in donors 1, 2, 3, and 4. Therefore, J\textsubscript{6}b and J\textsubscript{6}c truly represent two different polymorphic loci.

Analysis of Rearrangement Sites at the D-J\textsubscript{\alpha} Junction. Fig. 5 highlights the 5' rearrangement sites of the J\textsubscript{\alpha} segments in VDJ joining clones. Rearrangement of the J\textsubscript{3} segment is most frequently observed at the 5' end (base 1 in the Fig. 5) of the J\textsubscript{\alpha} coding sequence. Rearrangements of the J\textsubscript{2} and J\textsubscript{5} segments are most frequently observed 2-5 bp downstream from the 5' end of the coding sequence, whereas the J\textsubscript{4} segments most frequently rearrange 4-9 bp downstream. Rearrangements of the J\textsubscript{6} segments are widely distributed 3-21 bp downstream.

Characteristics of D Gene Family Usage. Approximately 91% (100/111) of the sequences from Fig. 1 had significant homology to one of the presently known germline D genes (6, 19, 20). Fig. 6 aligns the D regions of the sequences in Fig. 1 to known germline D1 segments with the most homology. As seen, there is often extensive trimming of nucleotides at both the 5' and 3' ends of the D segment, which may make some of the homology assignment inaccurate. In many cases, the homology to a given family member is <90%, suggesting that these sequences are probably derived from other unknown members of the same family. Specifically for the D1 segments, 8 of 19 sequences have two base pair changes at the same position; i.e., C → T at position No. 12, and A → G at position No. 13. This variant form of the known germline sequence is observed in all six donors. The germline DN4 segment, which is the only other known member of DN family, does not match the variant at positions no. 12 and no. 13. These findings suggest the presence
| A | B | C | D | E |
|---|---|---|---|---|
| 1-0A | TCCANA | A | (AT) | TCCANA |
| 1-0B | TCCANA | D | (AT) | TCCANA |
| 1-0C | TCCANA | G | (AT) | TCCANA |
| 1-0D | TCCANA | TCCANA | C | TCCANA |
| 1-0E | TCCANA | TCCANA | TCCANA | TCCANA |
| 1-0F | TCCANA | TCCANA | TCCANA | TCCANA |

**408**

VDJ Joining in Human Peripheral Blood B Lymphocytes
Figure 1. Nucleotide sequences of the VDJ joinings from normal PBL of six healthy adult donors. Sequences from each donor are grouped and are lined up in increasing order of length. The first number in the first row represents the donor number, and the second number represents the length of the clone in base pairs. The length is calculated from the first base of the Sall site in the Vx primer in FR3 and ends at the last base of the Pnl site in the Jx primer. Sequences of the primer used in the PCR are not shown. The second row indicates whether the sequence is in or out of frame. Each sequence is subdivided into Vx, N, D, N, and Jx regions according to the results of computer analysis against published germline Vx, D, and Jx genes. Names of the germline D and Jx genes with maximum homology to the segments used in the VDJ joining are shown in parentheses in the appropriate rows. When a specific germline D gene could not be determined, (ns, nonspecified), the N-D-N sequence is presented as one stretch of sequences. Deletions from germline sequences are indicated as asterisks. VDJ joinings carrying Jx2 gene segments that were obtained using a Jx2-specific probe (see Materials and Methods) are shown at the bottom. Donors from which these sequences were derived are shown in the parenthesis after the Jx sequence. P nucleotides as described by Lafaille et al. (21) are shown by small letters when present as tetrameric palindromes. Underlined are the sequences that were used to generate two oligonucleotide probes used to determine the relative frequency of these specific CDR3 in the B cell population, as described in Table 4.
of a new D segment that belongs to the DN family. We can not rule out the possibility that some of the other aligned sequences are derived from related but not yet identified members of the same family. Table 3 shows the coding frame of the D segments used in those sequences that were found in frame. All three coding frames were used with some preferential use of specific coding frames in certain D genes. 5 of the six in-frame sequences carrying a DA segment used the second coding frame. All five in-frame sequences using a DM1 gene used the third coding frame. All four sequences using a D21/9 gene used the second reading frame. Sequences using the DK segments (DK1 and DK4) preferentially used the first and third coding frames.

Possibility of Unusual VDJ Joining. In 11 of 111 total sequences analyzed (10%), we could not find significant homology with a known D segment. In addition, several sequences have extraordinarily long N regions (>15 bp), including some with homology to a known D segment. We reexamined these sequences using less stringent criteria by allowing for homology against both DIR and D gene segments regardless of orientation. The following are possible unusual VDJ joinings. Fig. 7 shows examples for each of the following categories: (a) Vw-J1a joining without D. (b) Inverted D joining (Vw ← D-J1a joining). (c and d) Double D joining (Vw-D-D-J1a) and (Vw-D ← D-J1a). (e and f) Vw-DIR-
Table 1. JH Gene Usage in 99 Randomly Examined VDJ Joinings

| Donors | J1 | J2 | J3 | J4 | J5 | J6 | Total | Percent |
|--------|----|----|----|----|----|----|-------|---------|
| J1     | 0  | 1  | 0  | 0  | 0  | 0  | 1     | 1.0     |
| J2     | 0  | 0  | 0  | 0  | 0  | 0  | 0     | 0.0     |
| J3     | 0  | 3  | 3  | 2  | 1  | 0  | 9     | 9.1     |
| J4     | 8  | 14 | 4  | 11 | 7  | 8  | 52    | 52.5    |
| J5     | 2  | 5  | 4  | 2  | 1  | 1  | 15    | 15.2    |
| J6     | 5  | 5  | 2  | 2  | 3  | 5  | 22    | 22.2    |
| Total  | 15 | 28 | 13 | 17 | 12 | 14 | 99    | 100.0   |

Frequency of Specific CDR3 Sequences in the B Cell Population as an Indication of Clonal Heterogeneity. To determine whether the primers used in the PCR amplified a large number or a relatively discrete number of CDR3 regions, we tried to estimate the frequency of two specific CDR3 sequences in the PCR products amplified from the lymphocyte population of three donors. Two oligonucleotide probes (2-106A-DP and 2-118B-DP) were synthesized that were specific for the CDR3 of the randomly picked VDJ clones 2-106A and 2-118B from donor 2 (underlined in Fig. 1). Triplicate filters of M13 libraries containing amplified CDR3s from three normal lymphocyte samples (donor nos. 2, 3, and 4) were then screened with these probes. One set of filters was hybridized with the Jx2 probes to establish the number of the VDJ clones present in the M13 libraries. The second and third sets of filters were hybridized with the 2-106A-DP probe and 2-118B-DP probe, respectively, to establish the frequency of these two specific clones in the population. As shown in Table 2, we found one positive plaque for the 2-106A-DP probe in M13 libraries from donor no. 2 only, and not from the other two donors (nos. 3 and 4). No positive clones for the 2-118B-DP probe were found after screening 18,000-20,000 plaques in each donor's library. From these data, we estimate that in the peripheral blood of donor 2, the two CDR3 sequences from which the probes were derived are present at a frequency of no more than one in 20,000 different CDR3s.

Table 2. Frequency of VDJ Joinings Containing the Jx2 Segment in Peripheral Blood B Cells

| Donor | Jx2 clones* | Total VDJ joinings† | Frequency of Jx2 clones |
|-------|-------------|---------------------|-------------------------|
| 2     | 80          | 6,320               | 1.27                    |
| 3     | 59          | 6,041               | 0.98                    |
| 4     | 184         | 10,384              | 1.77                    |

* Number of recombinant clones containing CDR3 amplified sequences present in M13 phage libraries that hybridize to a Jx2-specific probe. † Number of clones positive for a mixture of Jx consensus probes (JxC probes) derived from Jx segment sequences 5' to the Jx primer used in the PCR amplification.

Figure 4. Sequences of Jx genes. Jx3a, Jx4a, Jx5a, and Jx6a were original sequences published by Ravetch et al. (18). Jx4b and Jx5b were described as polymorphisms in Jx4 and Jx5 genes by Schroeder et al. (23). Jx6c was first described by Ravetch as a variant from Jx6 found in one of the two recombined genes from IgM-expressing chronic lymphocytic leukemia cells (18). Jx3b also appeared in the paper by Schroeder et al. (23), although they did not mention it as a polymorphism. The Jx6b was noted in the work by Bird et al. (22) from rearranged IgH in leukemic lymphoblasts. The first group for each Jx gene is the comparison at the nucleotide level and the second group is the comparison at the peptide level. In this study, Jx3b, Jx4b, and Jx5b genes were used exclusively by all six donors while both Jx6b and Jx6c genes were used in the population. Consensus sequences between each version are shown as bars, while sequences that are not identical are indicated. Deletions are shown as circles. For each Jx segment, the 5' coding region begins at the first nucleotide shown. The 3' ends of the Jx segments are not shown. The number of CDR3 clones using a specific Jx is shown in the right column.
Figure 5. Analysis of the 5' junction of the Jx genes in VDJ rearrangements. Six germline Jx sequences are lined up from the beginning of the heptamer signal (underlined). Numbering starts at the first nucleotide of the coding sequence. The position of the 5' junction of the Jx genes observed in each VDJ sequence are plotted above the nucleotide as an asterisk. The first TG in each Jx gene is underlined. The 3' ends of the Jx segments are not shown.

The sites of Jx gene rearrangement have been proposed to occur primarily at TG nucleotide sequences within the Jx gene segments (23). This observation appears to be applicable to our results. In the J3 sequence, TG first appears at position no. 2. In J2, J5, and J4, TG first appears at nucleotide position nos. 5, 6, and 8, respectively. After rearrangements, most of the 5' borders of these Jx segments seem to be clustered upstream of or around the first TG. On the other hand, TG first appears at position 22 in the J6 sequence,

Table 3. Usage of D Coding Frame in CDR3 Peptides

| D genes | 1st | 2nd | 3rd | Unknown* | Total |
|---------|-----|-----|-----|----------|-------|
| XP4     | 1   | 2   | 4   | 0        | 7     |
| XP1     | 1   | 2   | 0   | 0        | 3     |
| XP1     | 5   | 4   | 1   | 0        | 10    |
| D21/9   | 0   | 4   | 0   | 0        | 4     |
| A1/4    | 1   | 5   | 0   | 0        | 6     |
| K4      | 2   | 0   | 1   | 1        | 4     |
| K1      | 0   | 1   | 3   | 2        | 6     |
| N4      | 2   | 1   | 0   | 0        | 3     |
| N1      | 5   | 4   | 2   | 1        | 12    |
| M1      | 0   | 0   | 5   | 0        | 5     |
| M2      | 1   | 0   | 0   | 0        | 1     |
| LR5     | 0   | 0   | 1   | 1        | 2     |
| LR4     | 0   | 2   | 0   | 1        | 3     |
| LR1     | 0   | 0   | 0   | 0        | 0     |
| LR2     | 0   | 1   | 0   | 1        | 2     |
| LR3     | 0   | 1   | 1   | 0        | 2     |
| HQ52    | 0   | 1   | 0   | 0        | 1     |
| NS†     | -   | -   | -   | -        | 11    |

* Frame is not known because of extensive mutations, or base deletions, however, enough homology is present to assign the sequence to a known D gene.
† The D gene cannot be identified.
‡ 75 in-frame sequences from 99 randomly picked clones and seven in-frame sequences obtained by screening with the Jx2 probe are included.

The sites of Jx gene rearrangement have been proposed to occur primarily at TG nucleotide sequences within the Jx gene segments (23). This observation appears to be applicable to our results. In the J3 sequence, TG first appears at position no. 2. In J2, J5, and J4, TG first appears at nucleotide position nos. 5, 6, and 8, respectively. After rearrangements, most of the 5' borders of these Jx segments seem to be clustered upstream of or around the first TG. On the other hand, TG first appears at position 22 in the J6 sequence,

Table 4. Frequency of a Specific CDR3 Sequence among a Polyclonal Population of Peripheral Blood B Cells

| Donors | No. of clones positive for 2-106A DP | No. of clones positive for 2-118B-DP | Total no. of CDR3 sequences screened* |
|--------|------------------------------------|-------------------------------------|-------------------------------------|
| 2      | 1                                  | 0                                   | 18,960                              |
| 3      | 0                                  | 0                                   | 18,026                              |
| 4      | 0                                  | 0                                   | 21,168                              |

* Number of clones in M13 libraries that were positive for a mixture of Jx consensus probes (JxC probes).
Figure 6. Alignment of the CDR3 sequences that have homology to known germline D genes. In the top line of each panel, the germline D coding sequence is presented. Nucleotides that are different to the germline D genes are shown, while one that are identical are shown by bars.
Figure 7. Examples of unusual VDJ joinings. (A) Three CDR3 sequences with V-D-J joinings without a D segment. (B) One sequence with an inverted D joining (V-D-J,). (C) Five sequences with double D joinings (V-D-D-J). (D) Two sequences with double D joinings with the second D inverted (V-D-D-J). (E and F) Three clones containing DIR and D segments (V-DIR-D-J and V-D-DIR-J). (G) Four clones containing DIR segments (including inverted orientation) instead of D segments (V-DIR-J and V-D-DIR-J). The inverted DIR sequence is shown as the reverse complementarity strand of the germline D sequences. The sequences from the peripheral blood of the six donors are underlined. The best homology to published germline D sequences is shown.

which may explain the widespread sites of rearrangement found in this segment.

There is very little documentation of the frequency of D gene usage in a polyclonal lymphocyte population from healthy adults. The total number of human D genes has been estimated to be ∼30, based on an organization of five repeats, each containing six D gene families (6). The D5 cluster has recently been described as duplicated within the V gene locus (19). Some D genes may remain undiscovered. It is most likely that the DN1-like sequences that we have identified, which have consistent base pair changes from known sequences, are derived from one of these uncharacterized D genes. Schroeder et al. (23) reported that the DHQ52 gene segment is most frequently used (8/15) in H chain transcripts from a 130-d human fetus. In the results presented here, given the restraints of maximizing homology, the DXP family appeared to be the most commonly found (29/111). Similarly, in 13 somatic D sequences published by Ichihara et al. (6), seven were assigned to the DXP family. Thus, these studies may indicate that the expression or selection of specific D gene segments differ according to the developmental stage of the individual. It is also possible that the number of functional
germline DXP family members could be higher than for other D family members through duplication.

As observed previously, these results indicate that, in general, all three reading frames are used in human D regions to make productive rearrangements (6). However, there may be segments (DA1, D21/9, DM1) where a single reading frame is preferentially used, as is usually found among murine D regions (24). Translation of the CDR3 sequences indicates that 71% of these rearrangements are in the correct reading frame to allow translation of a functional H chain. Previous analysis of Abelson murine leukemia virus-transformed murine B cells indicated that ~40% of the cells contained two V_r-D-J_h rearrangements, of which, one was productive and one was nonproductive (25). Since peripheral B cells would be expected to possess at least one productive rearrangement, this frequency of nonproductive V_r-D-J_h rearrangements would predict that 71% (100/140) of the CDR3 regions would contain an in-frame rearrangement. This predicted frequency is in close accord with the number observed.

We have shown that tetrameric palindromic nucleotides, whose formation have been originally described in detail by Lafaille et al. (21), are most frequently found at the V_r-N and N-D_h junctions in the human CDR3, while they are rare at the D-N and N-J_h borders. This finding follows the observation that nucleotides in the 3' end and of the D_h segments and the 5' end of the J_h segments appear to be frequently removed by exonuclease activity.

Finally, we have found that unusual rearrangements make up a small but notable amount (~10%) of the rearranged IgH population. DIR segments, as described by Ichihara et al. (6) as being D-like sequences with irregular spacer lengths between joining signals, which could be involved in DIR-D or D-DIR joining by inversion or deletion, appeared to be used in our population. Also, we have found examples of D-D fusions resulting in both direct and indirect (inverted) V_r-D-D-J_h recombination. The mechanism that can account for these D-D fusions has been elegantly described in the mouse by Meek et al. (26), and previously proposed as a mechanism for creating additional antibody diversity by Kurosawa and Tonegawa (3). It is not clear if the examples of apparent V_r-J_h joining result from deletions of D in D-J_h joinings or reflect extensive exonuclease modification of the D segment in V_r-D-J_h joinings.

This work was supported by grants CA-10815 and CA-47983 from the National Cancer Institute.

Address correspondence to Giovanni Rovera, The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104.

Received for publication 30 April 1990 and in revised form 30 October 1990.

References
1. Tonegawa, S. 1983. Somatic generation of antibody diversity. Nature (Lond.). 302:575.
2. Yancopoulos, G.D., and F.W. Alt. 1986. Regulation of the assembly and expression of variable region genes. Annu. Rev. Immunol. 4:339.
3. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. J. Exp. Med. 155:201.
4. Alt, F., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three D-J_fusions. Proc. Natl. Acad. Sci. USA. 79:4118.
5. French, D.L., R. Laskov, and M.D. Scharff. 1989. The role of somatic hypermutation in the generation of antibody diversity. Science (Wash. DC). 244:1152.
6. Ichihara, Y., H. Matsuoka, and Y. Kurosawa. 1988. Organization of human immunoglobulin heavy chain diversity gene loci. EMBO (Eur. Molec. Biol. Organ.). J. 7:1411.
7. Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science (Wash. DC). 239:487.
8. Mullis, K.B., and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol. 155:335.
9. Yamada, M., S. Hudson, O. Tournay, S. Bittenbender, S.S. Shane, B. Lange, Y. Tsujimoto, A.J. Caton, and G. Rovera. 1989. Detection of minimal disease in hematopoietic malignancies of the B-cell lineage by using third-complementarity-determining region (CDR-III)-specific probes. Proc. Natl. Acad. Sci. USA. 86:5123.
10. Rathbun, G.A., and P.W. Tucker. 1986. Conservation of sequences necessary for V gene recombination. In Evolution of the Immune Response. G. Kelsoe and D. Schultz, editors. Academic Press, New York. 75–115.
11. Boyum, A. 1968. Isolation of mononuclear cells and granulocytes from human blood. Scand. J. Clin. Lab Invest. 21(Suppl. 97) 77.
12. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A Laboratory Manual. 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
13. Seliger, J., K. Ballas, A. Merold, U. Kotschi, J. Lyons, F. Eisenbeiss, N.D. Sinha, and G.P. Talwar. 1986. New preparative methods in oligonucleotide chemistry and their application to gene synthesis. Chem. Soc. 26:569.
14. Kwok, S., D.H. Mack, D.N. Kellogg, N. McKinney, F. Falloona, and J.J. Sninsky. 1989. Application of the polymerase chain reaction to detection of human retroviruses. In Polymerase Chain Reaction. H.A. Erlich, R. Gibbs, and H.H. Kazaian, Jr., editors. Cold Spring Harbor Laboratory Press, Cold Spring
NY. 151–158.

15. Bucan, M., T. Ysang-feng, A.M. Colberg-Poley, D.J. Wölgemuth, J.L. Guenet, U. Franke, and H. Lebrach. 1986. Genetic and cytogenetic localization of the homeobox containing genes on mouse chromosome 7. EMBO (Eur. Mol. Biol. Organ.) J. 5:2899.

16. Hanahan, D.J. 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166:577.

17. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5463.

18. Ravetch, J.V., U. Siebenlist, S. Korsmeyer, T. Waldmann, and P. Leder. 1981. Structure of the human immunoglobulin μ locus: characterization of embryonic and rearranged J and D genes. Cell. 27:583.

19. Matsuda, F., E.K. Shin, Y. Hirabayashi, H. Nagaoka, M.C. Yoshida, S.Q. Zong, and T. Honjo. 1990. Organization of variable region segments of the human immunoglobulin heavy chain: duplication of the D5 cluster within the locus and interchromosomal translocation of variable region segments. EMBO (Eur. Mol. Biol. Organ.) J. 8:2501.

20. Buluwela, L., D.G. Albertson, P. Sherrington, P.H. Rabbitts, N. Spurr, and T.H. Rabbitts. 1988. The use of chromosomal translations to study human immunoglobulin gene organization: mapping DH segments within 35Kb of the Cμ gene and identification of a new DH locus. EMBO (Eur. Mol. Biol. Organ.) J. 7:2003.

21. Lafaille, J., A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa. 1989. Junctional sequences of T cell receptor γδ genes: implication for γδ T cell lineages and for a novel intermediate fo V-(D)-J joining. Cell. 59:859.

22. Bird, J., N. Galili, M. Link, D. Stites, and J. Sklar. 1988. Continuing rearrangement but absence of somatic hypermutation in immunoglobulin genes of human B cell precursor leukemia. J. Exp. Med. 168:229.

23. Schroeder, H.W., J.L. Hillson, and R.M. Perlmutter. 1987. Early restriction of the human antibody repertoire. Science (Wash. DC.). 238:791.

24. Kabat, E.A., T.T. Wu, M. Reid-Millor, H.M. Perry, and K.S. Gottesman. 1987. Sequences of Protein of Immunological Interest. U.S. Dept. Health and Human Services. Washington, D.C.

25. Alt, F.W., G.D. Yancopoulos, T.K. Blackwell, C. Wood, E. Thomas, M. Boss, R. Coffman, N. Rosenberg, S. Tonegawa, and D. Baltimore. 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO (Eur. Mol. Biol. Organ.) J. 3:1209.

26. Meek, K.D., C.A. Hasemann, and J.D. Capra. 1989. Novel rearrangements at the immunoglobulin D locus. Inversion and fusions add to IgH somatic diversity. J. Exp. Med. 170:39.