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 J_H segments, J_H4 was found most frequently (in 52.5% of the sequences), and J_H1 and J_H2 segments the least frequently (~1% of the clones). A polymorphic J_H6 gene with a one-codon deletion accompanied by a base change was present in two of six patients. Preferential breakpoints were found for J_H2, J_H3, J_H4, and J_H5, although the breakpoints of J_H6 were distributed more heterogenously.

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 DN families. 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_H) diversity (D), and junctional (J_H) segments in the rearranged IgH genes (1, 2). The hypervariability of this region is due to the combinatorial assortment of the many V_H, D, and J_H 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_H, D, and J_H 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_H segments and for the J_H 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_H gene segments and of the translational frame used in individual D family genes, but also extends

J. Exp. Med. © The Rockefeller University Press • 0022-1007/91/02/0395/13 $2.00
Volume 173 February 1991 395–407

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 J_H gene locus. Also, we provide further evidence that novel rearrangements (V_H-J_H, V_H-inverted D-J_H, V_H-DIR-J_H, V_H-D-D-J_H, V_H-D-inverted D-J_H, V_H-DIR-D-J_H, V_H-inverted DIR-D-J_H, and V_H-D-DIR-J_H) 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 SalI and PstI cloning sites, homologous to the V_H FR3 and the J_H genes, used for amplification of the VDJ joining region of the IgH, were 5′ CTGTCGACACGGCCGTGTATTACTG-3′ and 5′ AACTGCAGAGGAGACGGTGACC-3′, respectively. In some experiments, to exclude a possible bias in amplification of the J_H segments, primers specific for J_H2 and J_H3 were used. These J_H2 and J_H3 primers differed from the antisense universal J_4 primer by a G to A base substitution at positions 16 and 9 of the universal J_H primer, respectively. The J_H consensus (J_HC) probes, which were used to detect all VDJ rearrangements, were a mixture of four oligonucleotides derived from the sequences of J_H2, J_H3, J_H4, and J_H6 genes just 5′ to the J_H antisense primer used for PCR amplification. These oligonucleotides had the following sequences; J_H2: 5′-CTGGGGCCGTGGCACCCTGG-3′; J_H3: 5′-CTGGGGCCAAGGGACAATGG-3′; J_H4: 5′-CTGGGGCCAGGGAACCCTGG-3′; J_H6: 5′-TCTGGGGCAAAGGGACCACG-3′.

The J_H4 probe also hybridized to J_H1 and J_H5 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 Kwok et al. (14).

After phenol/chloroform extraction and ethanol precipitation, an aliquot of the PCR-amplified material was digested with both SalI 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 ^{32}P-labeled J_HC 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 J_HC 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 "seqed" program. Each sequence was checked for the presence, in the correct orientation, of the J_H and V_H 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 V_H primer in FR3 starts with VDTAVYY (C). (*b*) The amino acid sequence of the J_H 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 J_H 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 J_H segment, when both D and J_H 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 V_H, D, and J_H 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 DH5αF′ *E. coli*, and ~5–10 × 10^3 plaques per 15-cm plates were transferred to nitrocellulose filters to generate replica filters. Duplicate filters were screened separately with the J_HC 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 J_H2 segment usage in the population was determined by screening separately a large number of M13 clones containing VDJ joinings on duplicate filters with the J_HC probes that hybridize to all six J_H genes and with the J_H2-specific probe. Stringent hybridization conditions (washing the filters at 1°C below the melting temperature of the probe) were used with the J_H2 probe in order to avoid crosshybridization with other J_H segments. 12 of the plaques that hybridized to the J_H2 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_H to the 3′ end of J_H 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_H2 segments that have been isolated using a J_H2-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_H border, five at the 5′ D border, one at the 3′ D border, and one at the 5′ V_H 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_H2 segments (58%) were in-frame. The in-frame sequences from Fig. 1 are grouped according to the J_H 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 then 136 bp are in-frame, in contrast to the smaller sequences in which the majority are in-frame.

Characteristics of the J_H Regions Utilized in Adult B Lymphocytes. Table 1 shows the frequency of J_H 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_H segments are J_H4 (52.5%), J_H6 (22.2%), and J_H5 (15.2%). Only one VDJ clone carrying a J_H1 segment and no clones containing a J_H2 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_H2 and J_H3 segments due to one-base mismatch with these sequences. Three J primers 100% homologous to J_2, J_3, and J_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_2, J_3, and J_4. The data (not shown) indicate that no differences in the frequency of J_H2, J_H3, and J_H4 containing recombinant clones were observed when either the J_H universal or J_H3 or J_H2 primers were used in the PCR.

Table 2 shows the frequency of J_H2 gene usage as determined by screening an M13 library containing CDR3-amplified sequences with a J_H2-specific oligonucleotide probe (see Materials and Methods section). The calculated frequency of J_H2 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_H3, J_H4, J_H5, and J_H6 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_H3b differs from J_H3a by one base (G to A), which results in an amino acid change from V to I. J_H5b differs from J_H5a 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_H6c 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_H6b. J_H4b and J_H6b differ from J_H4a and J_H6a by one base change, respectively, which does not result in an amino acid change.

J_H3b, J_H4b, and J_H5b segments were exclusively used by all six donors. For J_H3 and J_H5, 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_H3b, J_H4b, and J_H5b sequences should be considered as prototype sequences instead of J_H3a, J_H4a, and J_H5a sequences, which were originally reported as germline J_H sequences in a single individual (18). All the J_H6 sequences obtained from donor 6 used the J_H6c sequence. This variant sequence was also observed in two of three J_H6 sequences from donor 5, while J_H6b sequences were found exclusively in donors 1, 2, 3, and 4. Therefore, J_H6b and J_H6c truly represent two different polymorphic loci.

Analysis of Rearrangement Sites at the D-J_H Junction. Fig. 5 highlights the 5′ rearrangement sites of J_H segments in VDJ joining clones. Rearrangement of the J_H3 segment is most frequently observed at the 5′ end (base 1 in the Fig. 5) of the J_H coding sequence. Rearrangements of the J_H2 and J_H5 segments are most frequently observed 2–5 bp downstream from the 5′ end of the coding sequence, whereas the J_H4 segments most frequently rearrange 4–9 bp downstream. Rearrangements of the J_H6 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 DN1 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

CDR3 Clones	Frame	V	N	D	N	J	CDR3 Clones
1-88A	IN	TGCAAGA		(ns) TCCCACTGCGGATC		TTGACTACTGGGGCCAGGGAACCCT (J4)	1-88A
1-88B	IN	TGCGAGA		(XP'1) CGGAGT	GTCCC	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	1-88B
1-91	IN	TGCCAGA	AGA	(K4) GATTACAGCTATGG		TTGACTACTGGGGCCAGGGAACCCT (J4)	1-91
1-103	IN	TGCAAAAGA	TATGTTCGGA	(N1) GGTATGGC	TGGCAG	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	1-103
1-104	OUT	TACTAGGGA	G	(A1/4) GACTACTGCGACTAC	AACCACTGCG	TTTGACTACTGGGGCCAGGGAACCCT (J4)	1-104
1-106	IN	TGCGAGAGA	TTAGGATGG	(XP'1) CGTTCGGGGAGTTA	CCCCCT	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	1-106
1-112	IN	TGCGAGAGA	tcGATATGTACGCGA	(LR4) TAGTAGT*CCAG	AGGAGG	ACTTTGACTCCTGGGGCCAGGGAACCCG (J4)	1-112
1-115	IN	TGCAAGA	GTTAATCC	(XP1) GTATTACGATATTTTGACTG	ACA	ACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	1-115
1-118A	IN	TGCGAGA	AGG	(XP4) GTATTACGATTTTTGGAGTGGGTAT		CTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	1-118A
1-118B	IN	TACTGGA	TATACAC	(N1) ATAGCAGTGGCTGG	GTATC	CTACTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	1-118B
1-124	IN	TGCGAGA	CGGGC	(21/9) GTATTACTATGATAGTAGTAGTGGTTATTCCACC	AAGGG	CTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	1-124
1-127	OUT	TGCGAGAGA	TAATT	(K4) TACAGCTATGGT	CCCCCTACAACTAGTTCGACCCCT	ACAACTAGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	1-127
1-130	IN	TGCGAAA	CCAACGGAGTATAAA	(XP1) GTATTACGATATTTTGACTGGTTATTA	GGCGGGT	ACTACTTTGACTACTGGGGCCAGGGAACCCT (J4)	1-130
1-139	IN	TGCGAGAGG	ccCCCCACGTCC	(XP4) GTATTACGATTTTTGGAGTGGTTACCATACC	GTGG	ACTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	1-139
1-151	OUT	TGCGAGAGA	tcCTCATTTTGCTAAAACT	(N1) GTATAGCAGCAGCTGGTAC	AGCCCCCACAGGTAAAGGGGACTTGAT	AAACTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	1-151
2-94A	IN	TGCGAAAGA	A	(N4) GAGTATAGCACC	G	ACTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-94A
2-94B	IN	TGCGAGA	TCCAATTac	(K4) GTGGACCACAGCTACGGT		TGGGGCCAAGGGACAAT (J3)	2-94B
2-100A	IN	TGCGAAAGG		(ns) ATTCCACTCCAACGACGAGCGC		TTGAAATCTGGGGCCAAGGGACAGT (J3)	2-100A
2-100B	IN	TGCGAGGTA		(ns) TCCCGGCCT		CTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	2-100B
2-102A	OUT	TAAGAAAGA	TTCGGGGCC	(XP'1) ATTACTATGATACGAG	TGAC	GCACTGGGGCCAGGGCACCCT (J1)	2-102A
2-102B	OUT	TGCGAGAGG	TACCCC	(M1) TAACTGGAAC	GGGGTATGGG	GACTACTGGGGCCAGGGAACCCT (J4)	2-102B
2-103A	IN	TGCGAGAGA	TGG	(21/9) TTACTATGATAGTAGTGGTTA	CC	TTGACTACTGGGGCCAGGGAACCCT (J4)	2-103A
2-103B	IN	TGCGAGAGA	GGGGTA	(A1/4) TGACTACGGTGACTAC	GTTC	TTGACTACTGGGGCCAGGGAACCCT (J4)	2-103B
2-103C	IN	TGCGAGAGA	TGGA	(K1) GGACGTAGAGMTGGCTACAATT	GG	GACTACTGGGGCCAGGGAACCCT (J4)	2-103C
2-103D	IN	TGCGAGACA	tgAAGTC	(HQ52) CTGGGGA	TCCTG	AACTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	2-103D
2-106A	IN	TGCCAGA	G	(K1) ATAGT*GCCCCGAT	CGACTCCCCACGG	TACTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-106A
2-106B	IN	TGCGAAAGA	AGAGG	(XP4) CGATTTTTGGAGTGGTTATT	CCCTTCGG	TACTGGGGCCAGGGAACCCT (J4)	2-106B
2-109A	IN	TGCGAGA	GGGGGTTTTGGG	(A1/4) GACTACGTGACCTCC	TCACA	TTTGACTACTGGGGCCAGGGAACCCT (J4)	2-109A
2-109B	IN	TGCGAGA	GGGACCTACGGTGAC	(XP'1) TTCGGGAGCCCAT	ACTC	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-109B
2-109C	IN	TGCGAAA	AGT	(K1) CAACGATT	CACCCCAACAAAAAACGG	CGGCATGGACGTCTGGGGCCAAGGGACCAC (J6)	2-109C
2-109D	IN	TGCGAGAGA	GAGGac	(21/9) GTATTACTATGATAGTAGTGGTTATTA	CTACGTT	TGGGGCCAGGGAACCCT (J5)	2-109D
2-111	OUT	TGCGAGAGA	ATTGGGCCCGATTGC	(LR2) GGTATGATAGTAGTGG	C	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-111
2-115A	IN	TGCGAAAGA	tcCCCGCCA	(21/9) GCATTACTATGATAGTAGTGGTTATCA	CCG	TGACTACTGGGGCCAGGGAACCCT (J4)	2-115A
2-115B	IN	TGCGAGAGA	ACCTCGGGTCTCTC	(K4) ATACAGCTATGGT	CCAGG	ACGGTATGGACGTCCGGGGCCAAGGGACCAC (J6)	2-115B
2-115C	IN	TGCCCGA	ACCTT	(XP1) GTATTACGATATGTTGACTGGTTAT	ACCCCCGT	TTTTGATATCTGGGGCCAAGGGCAAT (J3)	2-115C
2-118A	IN	TGCAAGAGA	CGGTGGGGGA	(XP4) ATTTTTGGAGTGGTTA	TTACTCTG	AACTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	2-118A
2-118B	IN	TGCGAGAGA	CCCCGtc	(N4) GAGTATAGCAGCTCGT	TA	TACTACTACTACGGTATGGACGTATGGGCCAAGGGACCAC (J6)	2-118B
2-118C	IN	TGCCAGA	CATACTCTAACTACC	(N1) ATAGCAGCAGCTGGTA	TCGATCATGGG	TTTGACTACTGGGGCCAGGGAACCCT (J4)	2-118C
2-120A	OUT	TGTGAGGGG	GATGCGGTATTAC	(LR2) TATGATAGTAGTGGTTATTACTA	GAGAC	TTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-120A
2-120B	OUT	TGCGAAAGA	TTATGAC	(N1) GCAGTGGCTGGTTC	TTCC	ACTACTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	2-120B
2-121A	OUT	TGCGAAAGA	CCCGGGTAG	(21/9) TTACTATGATAGTAGTGGTTATTAGA	GGGG	CTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	2-121A
2-121B	IN	TGCGGTC	AACGGTG	(K4) GGGGATACAGTTATGGATA	TTTGGGGAACGGG	AACTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	2-121B
2-122	OUT	TGCAAGA	GCCATGGGG	(LR4) GGATATTGTAGTAGTACCAGCTGCTA	GTGCCCAGGC	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	2-122
3-79	IN	TACTAGA		(ns) GGGGTA		GACCCCTGGGGCCAGGGAACCCT (J5)	3-79
3-84	OUT	TACCACA		(ns) GT		TACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	3-84
3-94A	IN	TGTG		(N1) GCAGCAGGTGG	CCCAG	ATGCTTTTGATATCTGGGGCCAAGGGACAAT (J3)	3-94A
3-94B	IN	TGCGAGA	GGACCCC	(M2) CCGGAATCCC	TATTAACG	CCTGGGGCCAGGGAACCCT (J5)	3-94B
3-97	IN	TGCAAAGA		(N4) GATTATAGCAG	TGGCTGGT	TTTGACTACTGGGGCCAGGGAACCCT (J4)	3-97
3-100	IN	TGCGAGA	CAGAATTAC	(LR4) TATGATAGTAGT	GGTTGG	GACTACTGGGGCCAGGGAACCCT (J4)	3-100
3-103A	IN	TGCGAAAGA	TGGGG	(XP4) CGATT		TACTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	3-103A
3-103B	IN	TGCGAGA		(M1) TATAACTGGAACGAC	GAGATGAAGAAT	TTTGACTACTGGGGCCAGGGAACCCT (J4)	3-103B
3-109A	IN	TGCGAGAGA		(ns) GTCAAGGTGGAGCCCTGACCGACCGGGGGTA		TTTGATATCTGGGGCCAAGGGACAAT (J3)	3-109A
3-109B	IN	TGCGAAA	ACC	(N1) GGATATAGCAGCAGCTGG	ACCAGTGATGGTTC	TGATATCTGGGGCCAAGGGACAAT (J3)	3-109B
3-111	OUT	TGCGAGAGA	TGCCCTCGGGGGG	(N1) GCAGCAGCTGGTAC	GGG	CTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	3-111
3-118	IN	TGCGAGAGA	TGAAGGA	(XP'1) TACTATGGTTCGGGGA	AAGAGCCGGGGG	CTACTTTGACTACTGGGGCCAGGGAACCCT (J4)	3-118
3-126	OUT	TGCGAGA	AGCGCACCCC	(K1) AGTGGCTGT	CCTCCCGGTCGTGTTCTGGAGAGT	CAACTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	3-126
4-88	IN	TGCGAGA	TAC	(XP'1) TTCGGGG	TCC	TTGACTACTGGGGCCAGGGAACCCT (J4)	4-88
4-91A	IN	TGCGAGAGA	CGTGGGAGATGGGA	(XP'1) GGGGAG	G	CTGGGGCCAGGGAACCCT (J4)	4-91A
4-91B	IN	TGCA	GG	(LR2) AGGATCGGGTAGTG	CCCCG	GACTACTGGGGCCCGGGAACCCT (J4)	4-91B
4-91C	IN	TGTGAGAGA	TG	(M1) AGAACTGGA	G	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	4-91C
4-94A	IN	TGCGAGAGA		(ns) TCGGGATAGAG	G	CTGGTTCGACCCCTGGGGCCAGGGAACCCT (J5)	4-94A
4-94B	IN	TGCGAGAGA	TGGGG	(N1) AGCAGCAGC	CACC	TGACTACTGGGGCCAGGGAACCCT (J4)	4-94B
4-97	IN	TGCGAAAGA	tcG	(N1) GCAGCAGCTGG	CCA	ACTTTAACTACTGGGGCCAGGGAACCCT (J4)	4-97
4-100A	IN	TGCGAGA	CAAAACTTACCAACT	(XP'1) TACTACTATGG		TGACTACTGGGGCCACCAACCCT (J4)	4-100A
4-100B	IN	TGCAACA	GCCCCCCC	(M1) TAACTGGAACTAC	CC	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	4-100B

B

4-103	IN	TGTGAGA		(ns)	CCCCATGAAGGGGACGTACACAAC		ACTTTGAGTACTGGGGCCAGGGAACCCT (J4)	4-103
4-106	IN	TGCGAGA	GTAAATCcc	(N1)	GGGTATAGCAGTGGCTGG	TAC	TTTGACTACTGGGGCCAGGGAACCCT (J4)	4-106
4-109	IN	TGCGAGA	CTACTC	(M1)	AACTGGATATCC	ATGGGGGGCGG	TGCTTTTGATATCTGGGGCCAAGGGACAAT (J3)	4-109
4-115A	IN	CGCGAGAGA	GAAGTAT	(LR3)	GCATATTGTGGGGATGACTGCTCT	C	ACTTCTTTGATCACTGGGGCCGGGAACCCT (J4)	4-115A
4-115B	IN	TGCGAGA	GTCATCTCCCCTac	(XP'1)	GTATTACTATGAT	AGTATCA	ATGCTTTTGATATCTGGGGCCAAGGGACAAT (J3)	4-115B
4-120	OUT	TGCGAAA	GCCC	(K4)	ATACAGCGAAGGATA	TTAATGTCGGGAGCGAATCTGACCCCAGGGTAG	CCTGGGGCCAGGGAACCCT (J5)	4-120
4-124	IN	TGCGAGA	CTGTCCCGATACGG	(A1/4)	TGACTACGGTACCT	GGGCCGTGA	ACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	4-124
4-133	IN	TGCGAGAGA	TGGCCAGTACCACCCCCGGGA	(K1)	TGGCTACGAT	ACCCCCA	ACTACTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	4-133
5-91	IN	TGCGAGC		(ns)	TCTAAGACGCGCA		ACTTTGACTACTGGGGCCAGGGAACCCT (J4)	5-91
5-94A	IN	TGCGAGAGA		(ns)	GGGGGGTGGCCCCGGTTTG		GACTACTGGGGCCAGGGAACCCT (J4)	5-94A
5-94B	OUT	TGCGAGAGA	GA	(LR1)	AGTGGGGCGGCTATA	A	GACTACTGGGGCCAGGGAACCCT (J4)	5-94B
5-103A	IN	TGCAACA		(LR2)	AGGATGATAGTAGTGGT	CTTGG	ATGCTTTTGATATCTGGGGCCAAGGGACAAT (J3)	5-103A
5-103B	IN	TGCGAAA	AGGGAGTCCTGG	(A1/4)	GACTACAGTAAC	ACA	TTTGACTACTGGGGCCAGGGAACCCT (J4)	5-103B
5-105	OUT	TGCAAGA	ATCCC	(N1)	GTATAGCAGC	T	CTACTACTACGGTATGGACGTCTGGGGCCAAGGGACCAC (J6)	5-105
5-109	IN	TGCAAGA	GTCG	(M1)	GGTATAGTGGGATCTAC	CCACCCC	ACTAC***ATGGACGTCTGGGGCAAAGGGACCAC (J6)	5-109
5-112	IN	TGCGAAA	AGTGGTGGTCCCC	(N1)	GGTATAGCAGTGGCTGGT		ACTACTTTGACTACTGGGGCCAGGGAACCCT (J4)	5-112
5-115A	IN	TGCGAGA	CT	(LR5)	GATCACGGTGTCAGTCTT	CA	CTACTACAACTAC***ATGGACGTCTGGGGCAAAGGGACCAC (J6)	5-115A
5-115B	IN	TGCGAGA	CGTTACGG	(XP4)	GTATTACGATTTTTGGAG	CCGCGAGGCGC	ACTTTGACTACTGGGGCCAGGGAACCCT (J4)	5-115B
5-121	IN	TGCCAGAGG	TTTTG	(LR4)	ATTGTACTAGGAGTAGTTGCCA	CCGAAAACT	AAACTGGTTCGACCCCTGGGGCCAGGGCACCCT (J5)	5-121
5-138	OUT	TGCGAGAGA	tcTTTGT	(LR4)	GATATTGTAGTAGTACCAGCTGCTATGC	GGGGCGGAGTATAGCAGCTCGCGG	CTTTGACTACTGGGGCCAGGGAACCCT (J4)	5-138
6-85	OUT	TGCGAGA	CGA	(A1/4)	TAACTAC	gtGC	GACTACTGGGGCCAGGGAACCCT (J4)	6-85
6-88	IN	TGCGAGAGA	T	(LR5)	TGGGAATCGAGC		GACTACTGGGGCCAGGGAACCCT (J4)	6-88
6-94	IN	TGCGCGG	CIG	(A1/4)	GACCGCGGTAACT	CTCGGCCGC	CTACTGGGGCAGGGGA***CT (J4)	6-94
6-97	IN	TGCGAGAGA	G	(XP'1)	GGATTAGTGTGGTTCGGGG	CT	GACCTCTGGGGCCAGGGAACCCT (J5)	6-97
6-100	IN	TACCACC	CCCTCCGCCTACGG	(A1/4)	TGACTAC	CTA	TTTGACTACTGGGGCCAGGGAACCCT (J4)	6-100
6-101	OUT	TGCGAGA		(ns)	CGTGGAACCCG		ACTACTACAACTAC***ATGGACGTCTGGGGCGAAGGGACCAC (J6)	6-101
6-102	OUT	TGCGAGAGA	ACACCG	(LR5)	TTGAAAACGT	ACG	ACTACTTTGACTACTGGGGCCAGGGAACCCT (J4)	6-102
6-103	OUT	TGCGAAAGA	tc	(A1/4)	GACTACGGTAACTTC	CTG	ACTACTTTGACTACTGGGGCCAGGGAACCCT (J4)	6-103
6-109	IN	TGCGAGAGG	GGA	(XP4)	ACGATTTCTGGAGTCGTTCTT		CTACTAC***ATGGACGTCTGGGGCAAAGGGACCGC (J6)	6-109
6-112	IN	TGCGAGAGA	TACGCCCGTCCCCCGAG	(LR3)	TTGTGGCTGGTTTTG	AAACGAGG	TACTGGGGCCAGGGAACCCT (J4)	6-112
6-114	OUT	TACTAGA	CAC	(N4)	ATAACAGCTC	CCACTCTTCCTACTTCC	ACTACTAC***ATGGACGTCTGGGGCAGAGGGACCAC (J6)	6-114
6-120	OUT	TGCGAGAGA	CCGCGA	(N4)	GAGTAT	CAGATCCTAAATTCC	TACTACTACTACTAC***ATGGACGTCTGGGGCAAAGGGGCCAC (J6)	6-120
6-127	IN	TGCGAGG	ACTGTTGG	(N1)	GGGTATAGCAGTGGCTG	GCCCTAATTAC	TACTACTACTACTAC***ATGGACGTCTGGGGCAAAGGGACCAC (J6)	6-127
6-148	OUT	TGAGAAA	TCCCGCCTGTTGCT	(XP'1)	GTATTACTATGGTGTCGGGGA	AGGGACCCACGGGAAGGCGGTTTTGATAGGGGTGGG	TTGACTACTGGGGCCAGGGAACCCT (J4)	6-148
J2-105	OUT	TGCGAGAAA	ACATTTC	(XP'1)	TATTAT	GTGTAGGA	TGGTACTTCGATCTCTGGGGCCGTGGCACCCT (3)	J2-105
J2-106A	IN	TGCGAGA	GTTAA	(N1)	GTATAGCAGCAGCTGG	ATGGGTGT	CTTCGATCTCTGGGGCCGTGGCACCCT (4)	J2-106A
J2-106B	IN	TGCGAGA	GGGTCCCTG	(N1)	GGCCAACCCAG	AGAG	TGGTACTTCGATCTCTGGGGCCGTGGCACCCT (4)	J2-106B
J2-106C	IN	TGCGAGA	GTTCCG	(K1)	GGCTACGGTGA	Aag	CTACTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (2)	J2-106C
J2-107	OUT	TGCGAGA	TGGAGCTCC	(N1)	TGGCAGTGGCTGG		TACTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (3)	J2-107
J2-109	IN	TGCGAGAGA	CAGACCG	(N1)	GCAGTGGCTGGTAC	CCC	CTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (2)	J2-109
J2-110	OUT	TGCGAGAGA	TACAC	(LR3)	AGTGCTATT	GATAGTGTAGGA	TGGTACTTCGATCTCTGGGGCCAGGGACCCT (3)	J2-110
J2-118	IN	TGCGAGG	CAACC	(XP'1)	ACTATGGTTCGGGGGAGTTAT	ATGGGGCTG	ACTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (4)	J2-118
J2-120A	OUT	TCCGAGA	AGACCCAATA	(N1)	GCTTTAGCAGTGG	CCGGTACCGGAGG	ACTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (4)	J2-120A
J2-120B	OUT	TCCGAGAGA		(ns)	ACCCAAGAAAGAGCCAATGACTCGGCCCGGAAGA		ACTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (2)	J2-120B
J2-121	IN	TCCGAGAGA	tcGGGGCCGAGA	(K1)	TGGCTACAATAA	TCCCCCTTTTCAG	TGGTACTTCGATCTCTGGGGCCGTGGCACCCT (3)	J2-121
J2-139	IN	TACCACAGA	TGCCCCCGATCCCTAT	(XP'1)	TATTACTATGGTTCGGGGAGTTCTAGTA	GTAACGGGGG	CTGGTACTTCGATCTCTGGGGCCGTGGCACCCT (3)	J2-139

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 SalI site in the V_H primer in FR3 and ends at the last base of the PstI site in the J_H 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 V_H, N, D, N, and J_H regions according to the results of computer analysis against published germline V_H, D, and J_H genes. Names of the germline D and J_H 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 J_H2 gene segments that were obtained using a J_H2-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 J_H 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.

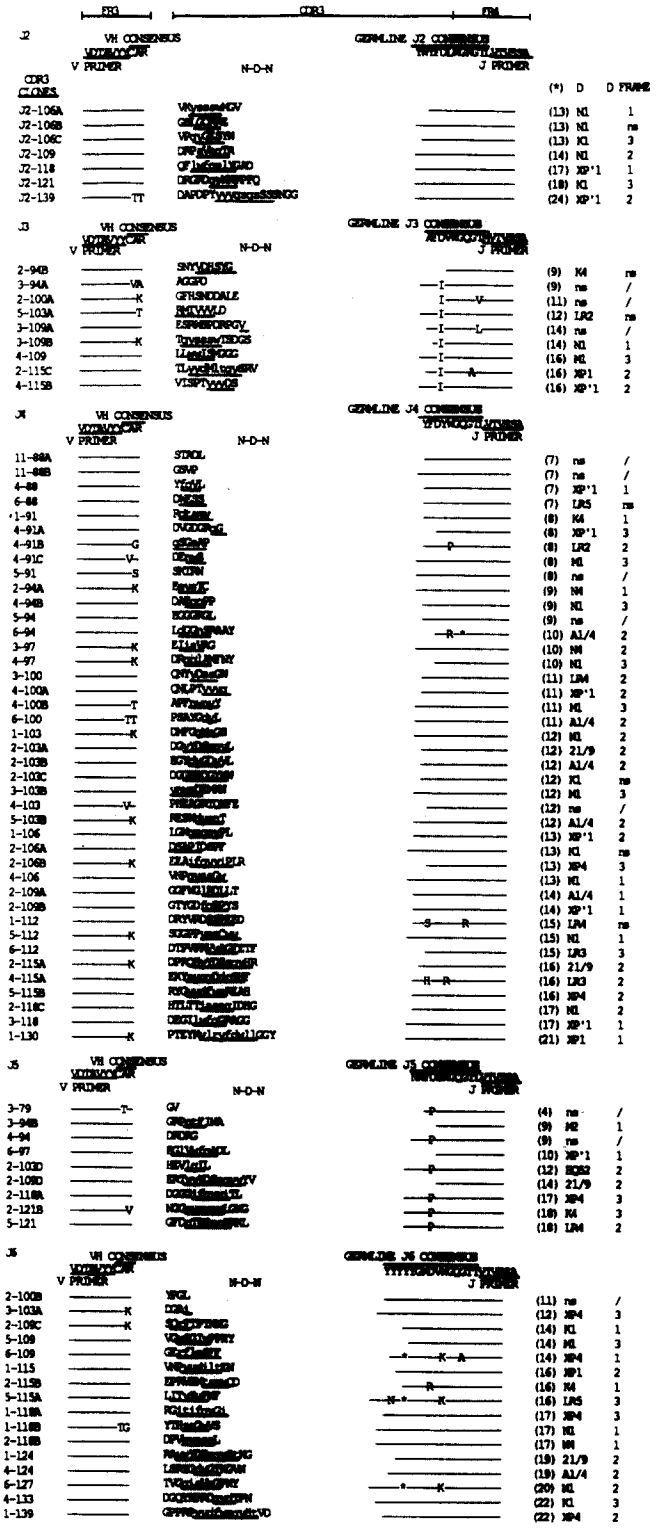


Figure 2. Translated amino acid sequences of the CDR3 and adjacent regions. 75 in-frame sequences from the 99 randomly picked VDJ clones and seven in-frame J_H2 sequences shown in Fig. 1 are included. The amino acid sequences are grouped according to the J_H segment used and lined up in increasing order of length. Amino acids derived from the V_H and J_H primer sequences may not entirely represent the original amino acids. The first 10 amino acids are lined up as the framework region 3 (FR3), amino acids identical to consensus sequences are shown by a bar, while amino acids that are not identical are indicated. The second block is the amino acid sequence derived from N-D-N region. Sequences homologous to germline D genes at the nucleotide level are underlined. Amino acids identical to those coded for by germline D genes are shown in small letters. The third block is the amino acid sequence derived from the germline J_H gene. Amino acids identical to consensus sequence are shown by bars, while amino acids that are not identical are indicated. The first column (*) after the sequences shows the length of CDR3 peptide. The second column shows the D gene used and the third column indicates the coding frame of the D gene used. (ns) The reading frame for a particular D gene cannot be identified due to either extensive deletion or mutation.

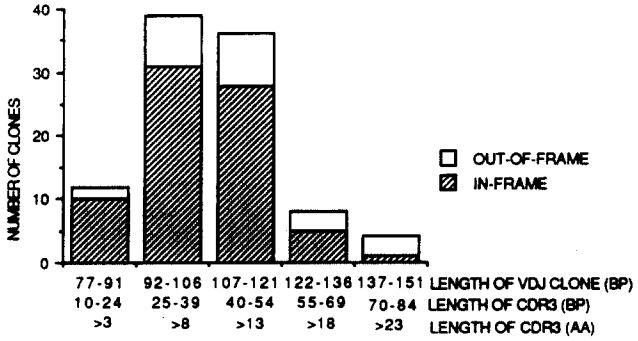


Figure 3. Length of the CDR3 sequences deduced from the 99 VDJ sequences from Fig. 1. Shaded bars represent in-frame sequences, while open bars represent out-of-frame sequences. On the *x*-axis, the first line shows the length of the VDJ joinings sequences in basepairs. These clones cover the 3′ end of the FR3, the CDR3, and the FR4. The second line shows the length of the CDR3 in basepairs. The third line shows the length of the CDR3 peptides (for in-frame sequences) in amino acids (aa). The *y*-axis shows the number of clones examined.

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*) V_H-J_H joining without D. (*b*) Inverted D joining ($V_H \leftarrow$ D-J_H joining). (*c* and *d*) Double D joining (V_H-D-D-J_H) and (V_H-D $\leftarrow$ D-J_H). (*e* and *f*) V_H-DIR-

Table 1. *JH Gene Usage in 99 Randomly Examined VDJ Joinings*

J_H segment	Donors 1	2	3	4	5	6	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

Table 2. *Frequency of VDJ Joinings Containing the J_H2 Segment in Peripheral Blood B Cells*

Donor	J_H2 clones*	Total VDJ joinings‡	Frequency of J_H2 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 J_H2-specific probe.
‡ Number of clones positive for a mixture of J_H consensus probes (J_HC probes) derived from J_H segment sequences 5′ to the J_H primer used in the PCR amplification.

D-J_H joining, V_H-D-DIR-J_H joining. (*g*) V_H-DIR-J_H joining, V_H-inverted DIR-J_H joining.

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 J_HC 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 4, 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

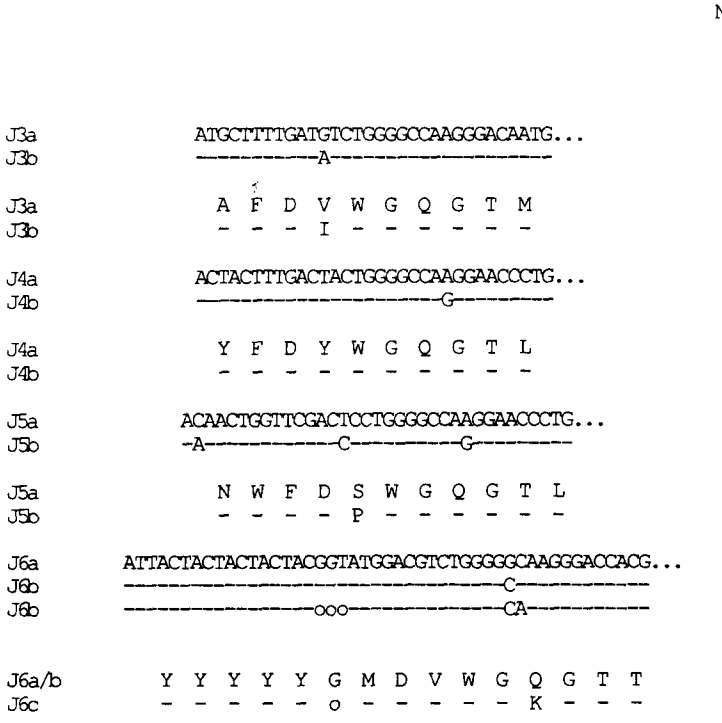


Figure 4. Sequences of J_H genes. J_H3a, J_H4a, J_H5a, and J_H6a were original sequences published by Ravetch et al. (18). J_H4b and J_H5b were described as polymorphisms in J_H4 and J_H5 genes by Schroeder et al. (23). J_H6c was first described by Ravetch as a variant from J_H6 found in one of the two recombined genes from IgM-expressing chronic lymphocytic leukemia cells (18). J_H3b also appeared in the paper by Schroeder et al. (23), although they did not mention it as a polymorphism. The J_H6b was noted in the work by Bird et al. (22) from rearranged IgH in leukemic lymphoblasts. The first group for each J_H gene is the comparison at the nucleotide level and the second group is the comparison at the peptide level. In this study, J_H3b, J_H4b, and J_H5b genes were used exclusively by all six donors while both J_H6b and J_H6c genes were used in the population. Consensus sequences between each version are shown as bar, while sequences that are not identical are indicated. Deletions are shown as circles. For each J_H segment, the 5′ coding region begins at the first nucleotide shown. The 3′ ends of the J_H segments are not shown. The number of CDR3 clones using a specific J_H is shown in the right column.

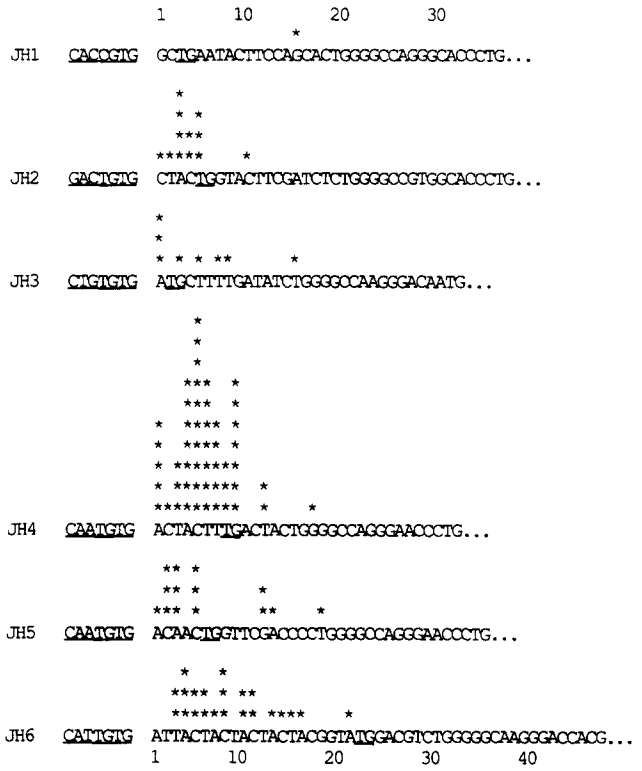


Figure 5. Analysis of the 5′ junction of the J_H genes in VDJ rearrangements. Six germline J_H sequences are lined up from the beginning of the heptamer signal (*underlined*). Numbering starts at the first nucleotide of the coding sequence. The position of the 5′ junction of the J_H genes observed in each VDJ sequence are plotted above the nucleotide as an asterisk. The first TG in each J_H gene is underlined. The 3′ ends of the J_H segments are not shown.

(representing 10,000–20,000 B lymphocytes, depending on the number of rearranged IgH alleles). These data also indicate that a single set of primers used for PCR amplification of VDJ joinings is able to amplify a very large proportion, if not the total representative population of the CDR3 region present in normal lymphocytes.

Discussion

We report a detailed analysis of the human H chain CDR3 present in the peripheral blood B lymphocytes of six healthy adults. To do so, we have amplified the CDR3 present in the cell population using the PCR with primers that are homologous to the 3′ end of the J_H segment (antisense primer) and to the FR3 of the V_H segments (sense primer) (9).

The data obtained indicate that J_H4 and J_H6 segments are the most commonly found J_H segments in adult peripheral lymphocytes of CDR3, and that J_H1 and J_H2 are the least frequently found. Schroeder et al. analyzed 15 IgH sequences in human fetal liver and reported a predominance of transcripts of J_H3 (seven clones), J_H4 (six clones), and J_H5 (two clones) (23). It is not clear at the present time if the differences between Schroeder's findings and ours are due to the small sampling in their study or different utilization or selection of J_H segments in a fetal vs. an adult B cell population.

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
XP′1	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
Total§	20	28	16	7	82

* 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 fom 99 randomly picked clones and seven in-frame sequences obtained by screening with the J_H2 probe are included.

The sites of J_H gene rearrangement have been proposed to occur primarily at TG nucleotide sequences within the J_H gene segments (23). This observation appears to be applicable to our results. In the J_H3 sequence, TG first appears at position no. 2. In J_H2, J_H5, and J_H4, TG first appears at nucleotide position nos. 5, 6, and 8, respectively. After rearrangements, most of the 5′ borders of these J_H segments seem to be clustered upstream of or around the first TG. On the other hand, TG first appears at position 22 in the J_H6 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 J_H consensus probes (J_HC probes).

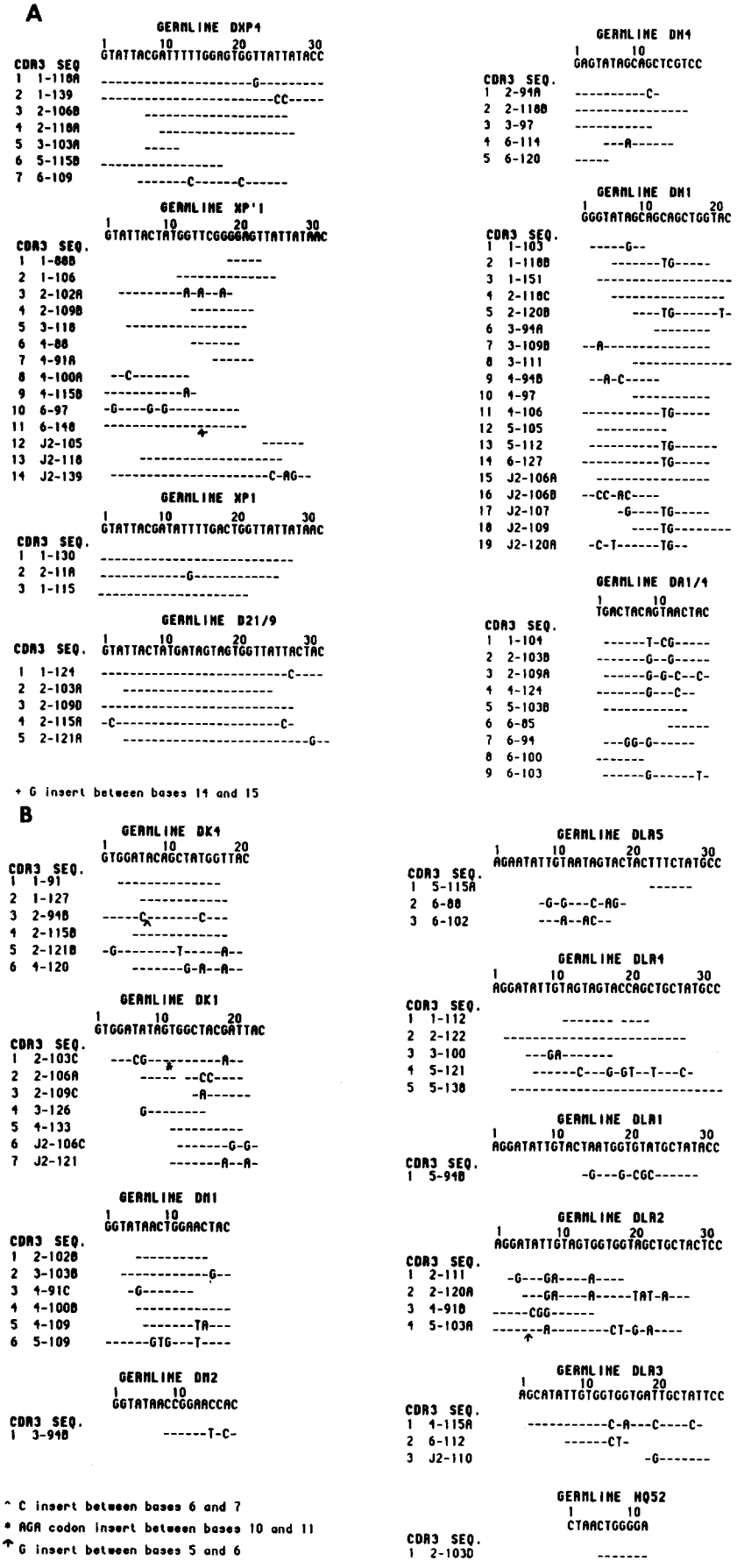
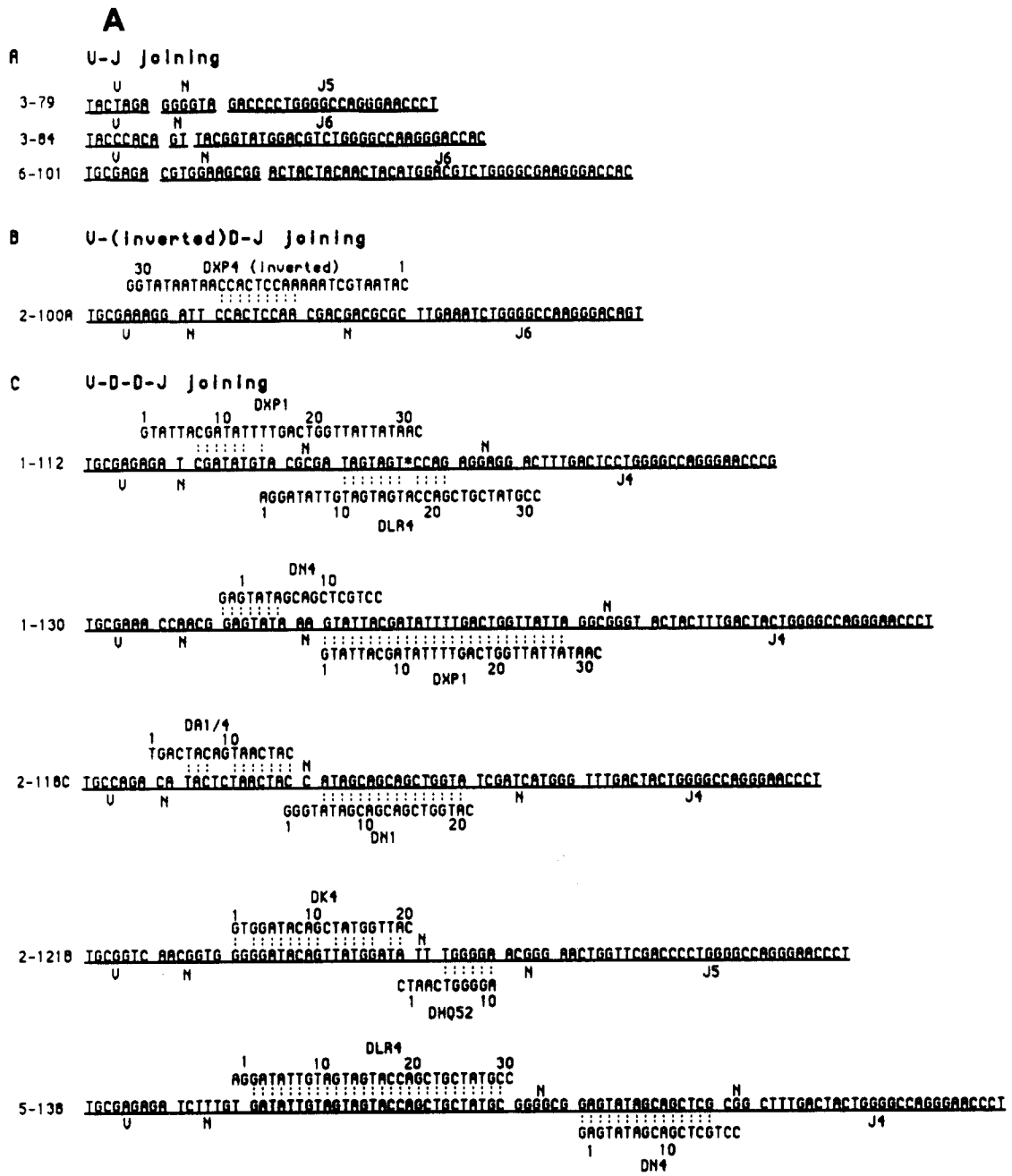


Figure 6. Alignment of the CDR3 sequences that have homology to known germline D genes. In the top line of each panel, the germline D coding sequence is presented. Nucleotides that are different to the germline D genes are shown, while one that are identical are shown by bars.



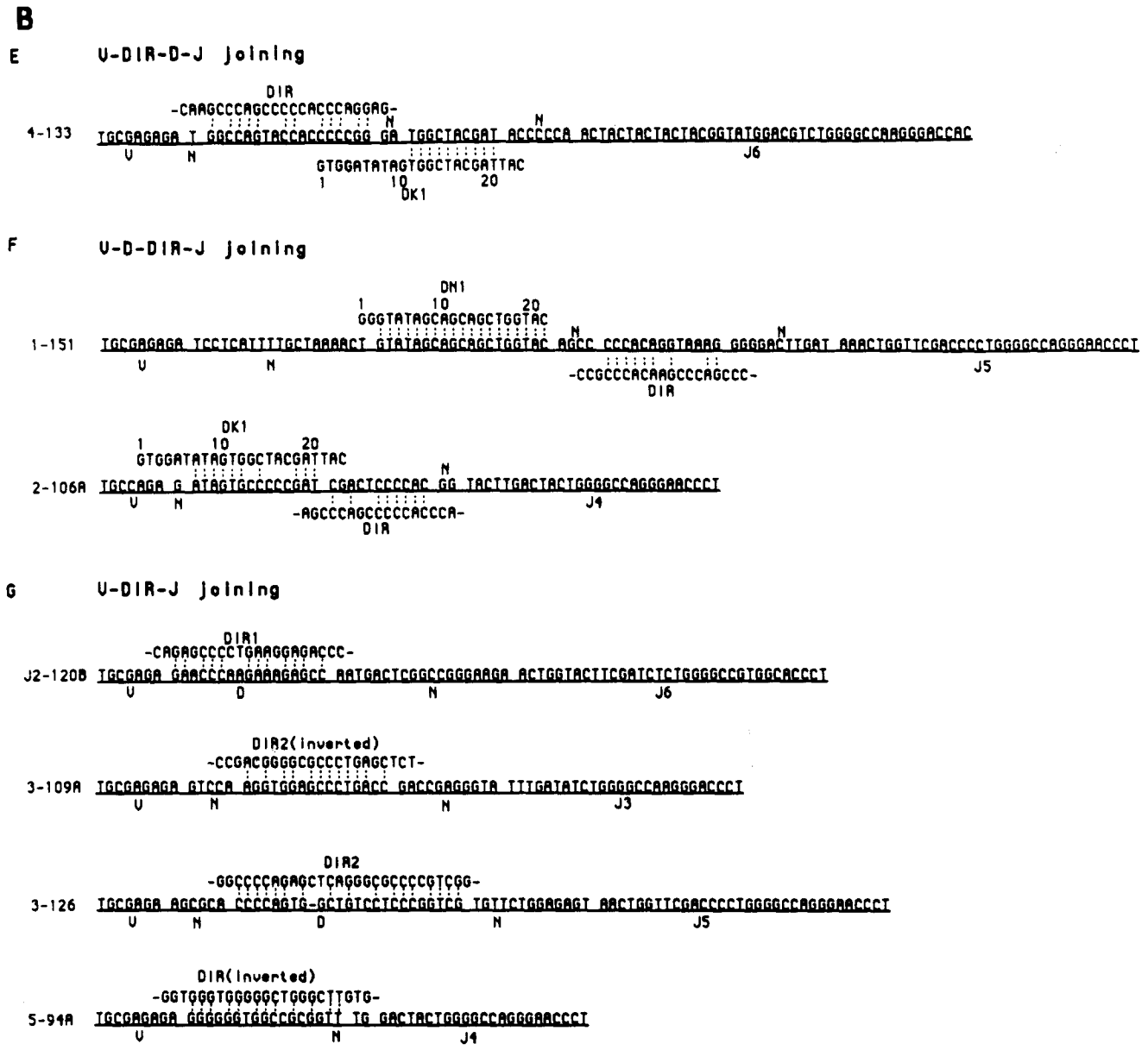


Figure 7. Examples of unusual VDJ joinings. (*A*) Three CDR3 sequences with V_H-J_H joinings without a D segment. (*B*) One sequence with an inverted D joining ($V_H \leftarrow$ D-J_H). (*C*) Five sequences with double D joinings (V_H-D-D-J_H). (*D*) Two sequences with double D joinings with the second D inverted (V_H-D $\leftarrow$ D-J_H). (*E* and *F*) Three clones containing DIR and D segments (V_H-DIR-D-J_H and V_H-D-DIR-J_H). (*G*) Four clones containing DIR segments (including inverted orientation) instead of D segments (V_H-DIR-J_H and $V_H \leftarrow$ DIR-J_H). 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_H 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 ~75% 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_H-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_H-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_H-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_H-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_H-J_H joining result from deletions of D in D-J_H joinings or reflect extensive exonuclease modification of the D segment in V_H-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_H 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. Mol. Biol. Organ.) J.* 7:4141.
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. Scr.* 26:569.
14. Kwok, S., D.H. Mack, D.N. Kellogg, N. McKinney, F. Faloona, 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. Kazazian, Jr., editors. Cold Spring Harbor Laboratory Press, Cold Spring

NY. 151–158.
15. Bucan, M., T. Ysang-feng, A.M. Colberg-Poley, D.J. Wolgemuth, J.L. Guenet, U. Franke, and H. Lebrach. 1986. Genetic and cytogenic 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 *Eschericha 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 D_5 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. Rabitts. 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 $\gamma\delta$ genes: implication for $\gamma\delta$ 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 repetoire. *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.