Immunology 2001 **103** 179–187

Third complementarity-determining region of mutated V_H immunoglobulin genes contains shorter V, D, J, P, and N components than non-mutated genes

KARLI ROSNER,*† DAVID B. WINTER,* ROBERT E. TARONE,‡ GUNHILD LANGE SKOVGAARD,†§ VILHELM A. BOHR* & PATRICIA J. GEARHART* **Laboratory of Molecular Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA, †Laboratory of Molecular Gerontology and Dermatology, Rigshospitalet, Copenhagen, Denmark, ‡Biostatistics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA and §Department of Dermatology, Bispebjerg Hospital, Copenhagen, Denmark*

SUMMARY

The third complementarity-determining region (CDR3) of immunoglobulin variable genes for the heavy chain (V_H) has been shown to be shorter in length in hypermutated antibodies than in non-hypermutated antibodies. To determine which components of CDR3 contribute to the shorter length, and if there is an effect of age on the length, we analysed 235 cDNA clones from human peripheral blood of V_H6 genes rearranged to immunoglobulin M (IgM) constant genes. There was similar use of diversity (D) and joining (J_H) gene segments between clones from young and old donors, and there was similar use of D segments among the mutated and non-mutated heavy chains. However, in the mutated heavy chains, there was increased use of shorter J_H4 segments and decreased use of longer J_H6 segments compared to the non-mutated proteins. The overall length of CDR3 did not change with age within the mutated and non-mutated categories, but was significantly shorter by three amino acids in the mutated clones compared to the non-mutated clones. Analyses of the individual components that comprise CDR3 indicated that they were all shorter in the mutated clones. Thus, there were more nucleotides deleted from the ends of V_H, D, and J_H gene segments, and fewer P and N nucleotides added. The results suggest that B cells bearing immunoglobulin receptors with shorter CDR3s have been selected for binding to antigen. A smaller CDR3 may allow room in the antibody binding pocket for antigen to interact with CDRs 1 and 2 as well, so that as the VDJ gene undergoes hypermutation, substitutions in all three CDRs can further contribute to the binding energy.

INTRODUCTION

The heavy and light chains of antibodies each contain three regions of hypervariability, termed complementarity-determining regions (CDR),[1] which interact with antigen. The most diverse of these is the third CDR of the heavy chain, which is located in the centre of the antibody binding site and makes more contacts with antigen than any other CDR. This region varies the most in length because it is constructed from several components. The heavy chain CDR3 is formed by amino acid residues encoded by a variable (V_H) gene segment, diversity (D) gene segment, and joining (J_H) gene segment. Using these multiple building blocks, further diversity is generated during joining by (a) addition of short palindromic (P) nucleotides to the ends of the coding sequences,[2] (b) deletion of a variable number of nucleotides from the ends of the coding segments by exonuclease activity, and (c) subsequent insertion of a variable number of non-templated (N) nucleotides at the V_H-D and D_H-J junctions by terminal deoxynucleotidyl transferase (TdT).[3] Additional diversity is introduced after joining by the hypermutational machinery, which introduces point mutations to change amino acid codons.[4] Thus in CDR3, both length and amino acid composition make major contributions to the antigen specificity. In contrast, CDRs 1 and 2 are relatively invariant in length and rely primarily on amino acid content to determine the binding affinity.

The length of CDR3 varies according to donor age and the hypermutation status of the V gene. Concerning age, a continuous increase in length occurs during fetal life until birth in mice and humans, which is primarily due to the relative absence of N regions in fetal genes.[5–8] Apparently this increase does not continue into adult life, as it has been reported that CDR3s from old people were the same size as those from young

Received 17 November 2000; revised 4 January 2001; accepted 23 January 2001.

Correspondence: Dr P. J. Gearhart, Laboratory of Molecular Genetics, National Institute on Aging, NIH, 5600 Nathan Shock Drive, Baltimore, MD 21224, USA. E-mail: gearhartp@grc.nia.nih.gov

adults.[9,10] However, since the cDNA libraries in these studies included genes with and without somatic mutations, a difference in length may become apparent if the regions are classified by mutation status. Concerning hypermutation, mutated antibodies have been shown to have shorter CDR3s than non-mutated antibodies in mice and humans.[11–13] In particular, Brezinschek *et al.*[12] found that the long J_H6 gene segment was used less frequently in mutated heavy chains than in non-mutated proteins, which contributes to the length difference.

To precisely determine the molecular basis of the shorter CDR3 length in mutated genes, we examined IgM transcripts of productively rearranged V_H6 genes from young and old donors. The lengths of the V_H, D, J_H, P and N elements were analysed to allow a comparison of the contribution of each to CDR3 diversity in mutated versus non-mutated antibodies, and to assess the impact of age on the length.

MATERIALS AND METHODS

Preparation of RNA from peripheral blood cells

Peripheral blood was collected from five young (26–29 years) and five old (81–86 years) participants in the Baltimore Longitudinal Study on Aging programme at the Gerontology Research Center, National Institute on Aging, National Institutes of Health in Baltimore, MD. The protocol was approved by the Institutional Review Board for Human Subjects Research of the Johns Hopkins Bayview Medical Center. Three people in the young group, Y1, Y3, and Y4 and four people in the old group, O1, O3, O4, and O5 were females. None of the participants expressed an acute illness at the time of blood removal. Twenty ml of peripheral blood was collected in ethylenediamene tetra-acetic acid (EDTA). Mononuclear cells were isolated by centrifugation through Ficoll-Paque Plus (Amersham Life Science Inc., Arlington Heights, IL), and total RNA was extracted using RNA STAT-60 (Tel.:Test B, Inc., Friendswood, TX).

CDNA cloning and sequencing

To make cDNA, 0·25 μg of RNA was transcribed with Superscript II reverse transcriptase (Life Technologies, Gaithersburg, MD) and a reverse primer complementary to the mRNA starting at codon 264 in the C_H2 exon of the constant (C) gene for IgM,[1] 5′AAGAAGCCGTCGCGGGGTGG. The cDNA was amplified in a 50-μl-reaction containing half of the cDNA, *Pfu* DNA polymerase (Stratagene, La Jolla, CA), a forward first primer for the leader region of the V_H6 gene starting at codon −19,[14] 5′TCTGTCTCCTTCCTCATCTTC, and the reverse first primer shown above. The amplification consisted of 30 cycles of denaturation at 95° for 1 min, annealing at 64° for 2 min, and extension at 72° for 3 min, followed by a final incubation at 72° for 10 min. Two μl of the reaction was then amplified for another 30 cycles using a second set of nested primers containing restriction sites for cloning. The forward second primer started at codon −10 in the leader region and contained a *Bam*HI site, 5′CGCGGATCCGCCCGTGCTGGGCCTCCCATG; and the reverse second primer started at codon 223 in C_H1 in the C_μ gene and contained a *Hin*dIII site, 5′TGGAAGCTTCACGTTCTTTTCTTTGTTGCC. The 670-bp PCR products containing both V and C genes were cloned into restriction-digested M13mp18. Viral DNA with rearranged V_H6 genes were sequenced with a primer starting at codon 140 in C_H1 of the C_μ gene, 5′AACGGCCACGCTGCTCGTATC.

Classification of cDNA clones and CDR3 components

Productively rearranged clones were classified as hypermutated if they had two or more mutations, and non-hypermutated if they had no or one mutation.[15] The CDR3 length was calculated by determining the number of nucleotides from residues 95 through 102.[1] The individual components of CDR3 were assigned as follows. (i) The V_H6 gene segment[14] contributed residue 95. (ii) D gene segments were identified if they were identical to germline sequences[16] for at least 8 nucleotides (nt), or if they had a single base substitution within a stretch of at least 9 nt. (iii) J_H gene segments[17] contributed codons up to and including residue 102. (iv) P nucleotides were identified in clones that had no deletions at the end of V_H6, D or J_H segments. (v) N nucleotides were identified as the bases at the V_H6-D and D-J_H junctions that could not be assigned to germline sequences or to P nucleotides.

Statistical methods

Comparisons of average gene segment lengths were performed using two-way analysis of variance methods, so that all comparisons between mutated and non-mutated clones were adjusted for donor age, and all comparisons between old and young donors were adjusted for clone mutation status. Pearson correlation coefficients were calculated to quantify associations between total CDR3 lengths and the lengths for specific gene segments. Comparisons of the usage distribution in the D and J_H gene segments were performed using Pearson's χ^2-square test. All *P*-values reported are two-sided.

RESULTS

The V_H6 gene was studied because it is the only member of its family,[14] it is non-polymorphic,[18] and it is present at similar levels in adult V_H repertoires.[19] This single V_H gene approach also circumvented possible bias in cDNA libraries due to preferential amplification of some members of specific V_H families.[20] V_H rearrangements to the C_μ constant gene were studied to include antibodies produced by both naive and memory B cells, and to obtain a more diverse library without potential restriction by a few dominant IgG clones. Some 235 unique productively rearranged cDNA clones were sequenced.[15] The mutated clones had an average frequency of 2·6% mutations per bp, which is within the normal range of mutations in IgM molecules from peripheral blood.[21] In this study, the data were analysed for gene segment usage and length of each component of CDR3 (Fig. 1).

Total CDR3 length

The CDR3 length distribution is shown in Fig. 2. The average length in the mutated clones was 33·0 nt, ranging from 21 to 60 nt, and the average length in the non-mutated clones was 40·6 nt, ranging from 15 to 72 nt. As shown in Table 1, this difference was significant ($P < 0{\cdot}0001$). The mean size of CDR3

was not different between clones from young and old individuals within the mutated or non-mutated categories ($P = 0{\cdot}55$).

V_H6 gene segment

The V_H6 gene segment makes a minor contribution to CDR3 of 0–2 nt, depending on exonuclease activity. As shown in Fig. 1 and summarized in Table 1, the average contribution was around 0·9 nt from mutated clones and 1·2 nt from non-mutated clones ($P = 0{\cdot}0023$). Thus, about 1·1 nt were deleted from the end of V_H6 in the mutated genes, and 0·8 nt was deleted in the non-mutated genes. Correlation of the V_H length to CDR3 length was significant in the mutated ($P = 0{\cdot}002$) and non-mutated ($P = 0{\cdot}035$) groups. There was no difference in length between clones from young and old donors within the mutated and non-mutated groups.

	Clone	V_H6 nt	P/N	D	D gene	P/N	J_H nt	J_H gene
Y1H-	5109	2	GGGGGAGCTACAA	AGTGGCTACGATT	D5-12	TGGGG	12	3
	5035	2	TGC	CTACGGTGAC	D4-17	AAG	12	4
	5062	0	GCT	TACGGTGGTAAC	D4-23	CGAGGG	9	4
	5122	2	CGGGGA	ATATAGTGGCTA	D5-12	TGTA	9	4
	5123	0		TCGACGACTGGGGCACGGGGG			9	4
	5070	1		GGCCTGGAATTCTC			6	4
	5102	2		AACTTTATCAGCGACTACTACACC			10	4
	5023	0		GGGGGACTGGTCCTCGGTACTCGGTCGGCAGAAGGGG			11	4
	5067	1	GGGC	GTtTAACTGGAtCTAC	D1-7	AATGGGC	20	6
Y2H-	J013	0		AGCAGTGGCTGGT	D6-19	CCCTCTCCCCACGAG	14	3
	J016	1	CTG	AGgGGGAGCTACTAC	D1-26	GG GG	13	3
	J008	2		GTATTACTATGATAGTAGTGGTTAT	D3-22	T	14	4
	J037	2	TTTGGGGCC	GAGTATAGCAGCTC	D6-6	CT	9	4
	J031	0	TA	TTTGtCTGG	D3-9	AGGGA	8	4
	J036	0		AGCCCTAACAATGGGCCTC			8	4
	J001	2	CC	CAACTGGA	D1-1	GATCTTTCTCTG	6	4
	J027	1		CACGATTCTACGGGA			11	5
	J017	2	T	AAGTCCG			26	6
Y3H-	4007	2	AGAC	ATAGTGGGAGCTACT	D1-26	CCCCCAC	14	3
	4073	0		ACCCGAGG			13	4
	4072	2	TGGCGGGGGCACC	GTGGCTAaGATTA	D5-12	TG	9	4
	4071	2	GGAC	ACTcTGGTTCGGGGAGTTATTA	D3-10	ATCCCC	8	4
	4158	0		TCGGTAGGATGGA			8	4
	4136	0		GGCCCCTCGCCCCTAAAAA			5	4
	4123	2	ACCCGAC	AGCAGCAG	D6-13	GGGA	15	5
	4001	2	CG	GTGGCTGGTA	D6-19		7	5
	4086	2	CTCAATCCTGGTCGGCTCCCAAAGAG	TGcTTCGGGGAG	D3-10	GT	12	5
	4112	1	GAAGAGAA	ATAGCAGCAGCTG	D6-13	ATCCAGTCCT	28	6
Y4H-	7177	0		ACAGTTAGACCCTTCACTACCAATG			14	3
	7107	1	GCCGGT	TAGCAGTGGCTGG	D6-19	GTGGAAC	12	3
	7182	2	TA	ATcTAGTGGCTACGATTA	D5-12	AG	12	3
	7155	1		GGGGCTTAAACGG			10	3
	7011	0	TCGGGTGGCGAGGGGATATGGGAGTGGG	AGTGGGAGCcAC	D1-26	GGT	14	4
	7183	1	GCGCCCGTG	GCAGCAGCTGGcAC	D6-13	CTCC	8	4
	7150	1	TT	TATTACGATTaTT	D3-3	CCCAAAG	10	4
	7004	0	ACA	AACTGGGGA	D7-27	AATACAGCTA	8	4
	7113	2	GGGGAGGAGTGGA	TGGTGGTA	D2-15		10	4
	7124	1		GTGTGAGAGGCAGCACGGCGAG			10	4
	7160	0		ACGGCAATCCGACGG			9	4
	7168	0		ACCCAGGGCTGGGCCACTCGAC			2	4
	7149	2		CGACCGGCATAGAATCAGCCTTC			8	6
Y5H-	2	0		AGGACAGGTGGCAGTAACG			14	3
	N81	1		GAATAGTAAACGC			10	4
	N66	0	ACGGA	AGTGGCTG	D6-19	TG	6	4
	N86	2	T	GGGGAAACTGGCCCAGAA			12	4
	N15	1		GTGGGAGCTAC	D1-26	CCCC	8	4
	N78	0		GGAGCGGGGCAGC			8	4
	N75	0	TCGAGCCC	GTATAGgAGCTCGTCC	D6-6	GGGGGGATC	6	4
	N90	0		TCTACACGTGGGGCTATG			6	4
	10	1		CTTTTCTTCTCAATCTGGGGCC			16	5
	L79	0		AGCCGTATCGGG			12	5
	L53	1		GGTGGCTCAGGGCAGGCG			11	5
O1H-	005	0	GTACCGA	ATAGCAGCAGCTGG	D6-13		12	4
	050	0	GTTGC	TAGAGATGGC	D5-24	T	14	4
	066	2		GGATTGGGATACTGGCTGGGT			1	4
	082	2	T	GGTAGAATATCACCTCGTCTTTCAT			14	5
	033	2	T	CGAGAGGAGTTTGCAGGGG			11	5
O2H-	021	1		GCCAAAGGGG			13	3
	025	0	GGAACTGCCACCTCAGC	TAcTAGTGGTTA	D3-22	CCCCCTTAGGG	2	4
	180	0		AGTCGGGGG			12	4
	037	2	AGGGG	GCAGTGGCTGGT	D6-19	CTG	11	4
	167	0		GGGTACCGGGG			10	4
	001	0	TGGTTGGA	TGGTAGttG	D2-15	GTGT	9	4
	013	1	GAGG	AGaAGTGGCTGGTA	D6-19	GATTA	9	4
	034	0		TCGCTGGATGGGAGTTCCCGA			9	4
	191	1	GGAGGCC	TATAGCAGTGtCTGG	D6-19	AGGA	9	4
	106	2	TTCGTCTAATCGTATCTCGCGTG	GCAGTGGCTGGT	D6-19	T	16	5
	088	0		AGTGGGTACTTCAAGGAATGG			12	5
	009	0		TGTATGCaATA	D2-8	TGG	13	5
	010	1		TTGCGGCAGAAAGCTCGTTCG			23	6
	111	1	CTGGCG	TAACTGGgACG	D1-20	GCCCTGTAT	21	6
	004	0	ATGCGATTGGGAATGAC	GAGATGGC	D5-24	CCGC	19	6
O3H-	6018	1		GGCCCGTAC			11	1
	6076	1		GACAGTGGCACCTC			12	1
	6052	2	TCAC	TATAGTGGaAGC	D1-26	C	14	4
	6066	2	CAAAGTA	GGTATAGCAGTGGCTGGTAC	D6-19	GGTGAGAT	14	4
	6154	1		GCCCCGAAC			11	4
	6079	1	GGGGGC	ATGGTGTA	D2-8	ACAGTGTCTCT	10	4
	6081	0	GG	AACTGGgAC	D1-1/D1-7/D1-20		10	4
	6074	0	CTCG	CAaCTGCTA	D2-2	CAGGCGCCG	8	4
	6010	0		TCCTCTGGCCCGC			8	4
	6077	0		AATAAGGATAGG			12	6
O4H-	8003	2	AAAGAGGGATTACTATGATACCACGGA	GTACAACTGGAACGAC	D1-1	CC	10	4
	8010	2		GTTGTCGT			14	4
	8043	1	T	TGGCTcGTA	D6-19	T	12	4
	8044	0		AGGAATGAGTGGGACTCT			9	4
	8004	0	GCC	AGCAGCAG	D6-13	G	12	4
	8039	2	A	TATAGCAGTGG	D6-19		10	4
	8050	0		GGCAAGAGAACTGG			10	4
	8007	1	CTACCCTGAATCA	GTtTAGCAGTGGCTGGTA	D6-19	TGGAG	11	5
	8033	2	TC	GACGACAGTGGCAAGG			25	6
	8024	0	GGG	GTAGTgGTACCA	D2-2	AGCCTTTCCTT	10	6

Figure 1. *Cont.*

	Clone	V_H6 nt	P/N	D	D gene	P/N	J_H nt	J_H gene
O5H-	9052	0	TTT	GTGGCTGttA	D6-19	TTG	14	3
	9067	1		TGGGGAGCTGGGCCGCCCACGCCCAGTTTG			14	3
	9009	0	GGGAACC	TGGGAGCTACT	D1-26	CTCCAGAGCTGCGGC	9	3
	9008	2	AACA	GATAGTAG	D3-22	CCCCT	14	4
	9020	0		TATAcCAGCAGCTGGTAC	D6-13	C	8	4
	9022	1		GGAGTaATCG	D3-16	G	12	4
	9011	0		TCCCCGGCAGCTCC			10	4
	9053	0		GGGTCGGCTGGCTATCTACC			10	4
	9074	0	TC	TAGTGGCaAC	D5-12	TCACTTCT	10	4
	9028	1	GT	ATGGTGTcTGC	D2-8	CCAGACT	9	4
	9041	2	TC	GGGGAAGGAGCTCCCGGCTGGGG			9	4
	9101	1		CAGTGGCTGGTA	D6-19	GGGGGGGG	9	4
	9029	2	T	GTGAGAGGGGGGAATCCGCAGC			8	4
	9099	1		TCGCGGCGGGTCTTGGG			12	5
	9072	1		CAAAGATCGGTGC			19	6
	9066	0	G	GTATAGCAGCTCGTCC	D6-6	GGGTCTG	27	6
Y1NH-	5009	1	CAAAG	GACTACGGTGGTAAC	D4-23	CGAA	17	2
	5033	2	TC	GTCCCTCGAGGGCAGTTTACGGGGGGGGAG			14	3
	5032	2	TC	ACTCCTACTTCTC			10	4
	5047	1	GG	CTGGAACTAC	D1-7	GTCTTTCACCT	9	4
	5085	1	GAGG	TAGCAGTaGCTGG	D6-19		9	4
	5069	0		GCACAGGTGGTTCTCAACAAGATCCCGGGA			15	5
	5053	2	TCCATTTT	TAGCAGCTCGT	D6-6	ACCC	26	6
Y2NH-	J034	1	TTGTGTTCATCGACC	ATTACTATGATAGTAGTGGTTAT	D3-22	CCGG	14	3
	J014	1	GATCGGTCG	GGTATAGTGGGAGCTAC	D1-26	CCGGG	13	3
	J026	2	ACCCCAA	GGGTATAGCAGCAGCTG	D6-13	GT	14	4
	J053	2	TCC	GGTACAACTGGAACG	D1-1	TGGGTAT	12	4
	J025	2	TACG	GCAGTGGCTGGTA	D6-19		11	4
	J021	0		TCGCATAGGAGTGC			10	4
	J015	2	GGCAATCA	TGACTACGGTGACT	D4-17	CACGGCCACTTTGGGC	8	4
	J019	2		ACACTGGTACGCCC			8	4
	J024	1	GTGGGGGCC	GATATTGTACTAATGGTG	D2-8	ATGCATACACCAACTA	7	4
	J032	2	AC	TcGTGGGAGCTA	D1-26	TCG	8	4
	J004	0		CTGAGAGGGAG			13	5
	J018	2	TCAAATCGAG	ATTACTATGGTTCGGGGA	D3-10		27	6
	J058	0	GGGGGGGT	GTATAGCAGCcCG	D6-6	GACTCATGGGGT	18	6
Y3NH-	4157	1	TCTGG	GCAGCAGCTGG	D6-13	CCACATTGGGAACTC	13	3
	4077	1	GGGGTTCGCCGGGGCAGCTC	GGcTCGGGG	D3-10	GTCA	8	4
	4070	1	GGTTG	GTGGGAGCTACT	D1-26	CGAGGA	24	6
	4094	2	T	GTTGCCGATTCGGGA			18	6
Y4NH-	7068	0		ACCCTATTGGG			13	3
	7179	0		AAATCCGGGGGACCC			9	3
	7111	2	GCGAGG	AGCAGCAGCTGGTAC	D6-13	GTGG	12	4
	7112	0		GGAAACGCA			6	4
	7006	2	AAA	GTATTACTATGATAGTAGTGGTTATT	D3-22	GGAG	25	6
	7074	0		GGgGGTGACT	D2-21	C	25	6
	7175	1	GG	GCtGTGGCTG	D6-19	ATCT	22	6
Y5NH-	7	2	TTCGTGGCTC	GTTCGGGGAGgTATT	D3-10	CCCGC	13	6
O1NH-	121	2	GG	TAGTGGGAGCTACTAC	D1-26	GTT	13	1
	127	2	GAG	TTGTGGTGGTGAtTGC	D2-21	CAAT	14	3
	081	2	TG	AGTGGCTGG	D6-19	ACCGGGAGT	14	4
	052	0	GCCT	TAGCAGCAGCaGG	D6-13	A	12	4
	132	2	GGGG	CTATGGTTCGGGGAGTTATT	D3-10	TGAGGGGCGGT	14	4
	008	0	AATTTCTTGTATGA	CAGCAGCTGG	D6-13	CCAAAGA	11	4
	037	2	TC	GATTAATATCACCTGGGGATAGGGG			10	4
	063	1	CA	TATTACTATGGTTCGGGGAGTTATTATAA	D3-10	TA	2	4
	109	0	GGC	CAGCAGCTGGTA	D6-13	TTTCGGGG	1	4
	071	1	GTGAGGGA	GACTACGGTGACTAC	D4-17	GGtGAGTTC	0	4
	011	1	GGATCGGG	GATGGCTACAA	D5-24	CCCGG	17	5
	028	0	CGGAA	AACTGGAACTAC	D1-7	GTCTG	14	5
	034	1	CTCCGGTGGGGGTTGGTAGAATGG	TTACTATGGTTCGGGG	D3-10	CTAC	27	6
	072	0	GGGGAG	GCAGTGGCTGGT	D6-19	CC	25	6
	014	1		GGGCTGGGGAAC			23	6
	113	2	TCGCT	GTAGTACCAGtTGC	D2-2	CCA	24	6
	060	0		GTACGCGATG			20	6
O2NH-	154	1	CGGACTTAAG	TTACTATGGTTCGG	D3-10	CACGAG	14	3
	090	2	TATTG	TAgTGTGGTGGTGA	D2-21	GCTGATGCG	9	3
	189	1	T	TATAGCAGTGGCTGGTAC	D6-19	TTTGGGGATCTGGGGC	3	3
	169	2		CTATGATAGTAGTGGTTAT	D3-22	AATTGGGAGGGG	12	4
	139	0	TCG	AcAGTGGCTACGATT	D5-12	CTCGTCCGTTCCTGGTTATTAGG	10	4
	142	2	GTG	GGATAcTGTAGTGGTGGTAGCTGC	D2-15	GTACACAGATCG	10	4
	194	2	T	AGCAGCAGCTGGTAC	D6-13	TCCCC	10	4
	003	2	AGGTGGGAGATC	GTATAGCAGTGGCTGGTAC	D6-19	GCGCCCCTACTA	9	4
	098	0	AACGTCCA	GTTCGGGGAG	D3-10		9	4
	060	1	TG	GTGGGAGCTAC	D1-26	GC	11	5
	095	2	A	ATAGTGGCTAC	D5-12	TCATCTG	9	5
	105	1	AAGTCGTCGAATCGGGGA	GGATACAGCTATGGT	D5-5	CCCAAAAAAT	28	6
	121	2	TCATTC	GTATAGCAGCAGCTGGT	D6-13	CTGACGCG	27	6
	196	0	TACTTTCAAATTCTGC	ATTACTATGGTTCGGGGAGT	D3-10	CGAA	23	6
	156	2	GCTCTGGCGGTCTGGTCGGGCTGCCG	GTGGGAGCTACT	D1-26	T	22	6
	144	1	TTACACC	TATAGCAGCAGCTGGTAC	D6-13	TG	26	6
	006	2	TCACGGG	ATAGCAGTGGCcGC	D6-19	CGGTGAGGAAG	23	6
	072	0	TTTTATCtTTTCGG	TACGGTGACTAC	D4-17	ATTTACGACCA	17	6
	128	2	TCTGGGGCGG	ATAGCAGCAGCTGGTA	D6-13	TTTTCGGGCCTAACTG	13	6
	035	0		CGGCATAGACTGGG			13	6
O3NH-	6040	2	T	GTATTCTGGGATCAAG			14	2
	6051	1	GG	AGTGGGAG	D1-26	GAGCCCGTTCG	14	3
	6128	1	GGCTAAGATC	GATAGTAGTGGTTATT	D3-22	TTGG	14	3
	6133	2	TACGG	GGATAgAGTGG	D5-12	TACCCGA	14	3
	6002	2		CCGGGTTGAGAGGTGG			12	3
	6063	1	CCCGTCC	TATTGTAGTAGTACCAG	D2-2	GCTCCCCGCAG	12	3
	6099	1		GTTCCGCACCCAGGGGGGCGGGGGAGGGG			12	3
	6042	2	CCGCACA	TATAGCAGCTCG	D6-6	AAC	12	3
	6118	2	TACA	TATAGCAGCTCGTCC	D6-6	TTAGTAGCCTT	10	3
	6157	2	TTCGGAGGtCCTTCA	CAGCAGCTGGTAC	D6-13	GGGA	14	4
	6083	2	AGATGGC	GATGGCTAC	D5-24	C	11	4
	6053	1	GG	TACGATTTTTGGAGTGGTTATT	D3-3	T	13	4
	6167	2	C	GGATATTGTAGTGGTGGTAGCTGCTcCT	D2-15	G	13	4
	6126	2	TGCG	GCAGCAGCTG	D6-13	CCCCCGGG	12	4
	6086	2	TTCCAT	CAGTGGCTGGTAC	D6-19	GATG	11	4
	6162	1		TTACGATTTTTGGAGTGGTT	D3-3	TCGATTACCG	11	4
	6139	2	TGGAACG	CAGCTGGGAC	D6-13	GTTGGGACCCAGGG	9	4
	6152	2	GGAAAAAG	GATTTTTGGAGTGGTTATTAT	D3-3	TTTTCTTGCTG	9	4
	6043	2	CAGAAAAGA	TGaATACAG	D5-5	AAGTG	8	4
	6127	2	GC	ACAACTGGAACGAC	D1-1	GGGTTCA	8	4
	6144	1	GGGTGG	GTGGCTGGTAC	D6-19	GTAA	8	4
	6125	2	TCGG	AGTGGCTG	D6-19	GT	17	5
	6170	2		TCTTTGGGCCAACCAGTGGCGGCCG			15	5
	6169	1	GT	TATAGCAGTGGC	D6-19	GGGG	14	5
	6056	0	GGTAGGTCA	AGCAGCAGCTGG	D6-13	CGGAGGGG	13	5
	6146	0		ACGG			29	6
	6087	2	GGGATACAGGG	GCAGCAGCTGGTA	D6-13	TCAAA	26	6
	6090	0		TTAGTAAGCTGTACCGCTTC			22	6
	6123	2	TCCACCCGCTG	TAGTGGGAGCTAC	D1-26	GA	20	6
	6062	0	TCTTATCC	GGGTATAGCAGCAGCTGGTA	D6-13	GGGATTTGAGGC	14	6

Figure 1. *Cont.*

	Clone	V_H6 nt	P/N	D	D gene	P/N	J_H nt	J_H gene
O4NH-	8011	1		GAAGCGGGACCAATGGAGAATCGTGGTACTCTG			14	3
	8021	1		GGGCCTGGGGTGGATCAAGGGG			13	3
	8062	0	TACGTCGGAACCAGA	AGTGGGAGtTAC	D1-26	GGGGTAATTGG	13	3
	8035	0		AAGAATGCGGGT			12	3
	8022	1	ATGAGGAGGGTTACAACCCCAGGGTGGAT	GGTATAGT	D1-26		10	3
	8009	2	ACCCCTAGAACT	GTATAGCAGCTC	D6-6	AGTCT	14	4
	8013	2	T	AAgTGGAACT	D1-7	C	10	4
	8048	1		GGgATAGCAGCAGCTGGTA	D6-13	T	9	4
	8001	2	AGA	AGCAGTGGCTGGTAC	D6-19	CCCT	9	4
	8031	2	TTGCCTGGCCGc	GTATAGTGGGAGCTACT	D1-26	GGC	8	4
	8036	1	GGGGTTcCAATATTG	GCAGTGGCTGGTAC	D6-19	TGG	6	4
	8012	1	CA	TATAGCAGTGGCTGGTAC	D6-19	CTGGGCT	11	5
	8053	2	TCT	TGGAGCTC			26	6
	8023	1	ATAGTGT	GTATTACGATTTTTGGAGTGGTTATTAT	D3-3	GGGAGGG	20	6
	8037	1		GCTGGTAC	D6-13/D6-19		9	6
O5NH-	9004	2	TCCG	GCATATTGTGGTGGTGACTGCTAT	D2-21	CCTCCACGAG	14	3
	9134	2	TACGGGGA	TAgAGTGGGA	D1-26	CCCGA	14	3
	9139	2	TCTCCTACT	AGGATATTGTAGTGGTGGTAGCTGCT	D2-15	CCACTGT	13	3
	9064	0	TTT	GTGGCTGtTA	D6-19	TTG	14	3
	9095	0		ATAGCAGCTCG	D6-6	GTCTT	14	4
	9049	1		TTTTGTTGAGGAGGGGGCTCACGGATGG			10	4
	9135	1	GCTTCCAC	GGGTATAGCAGTGGCTGGTAC	D6-19	CCC	6	4
	9013	1		CCAGGGGAGCTCGCA			14	5
	9051	2	CGtCAGTGCC	GGGTATAGCAGCAGCTGGTAC	D6-13	CTTAACTGG	15	5
	9015	0		AGCAGCAGCTGGTAC	D6-13	GTCGGG	12	5
	9037	2	GACTCAGGTTA CC	GGGTATAGCAGCGGCT	D6-25	GGTACCG	22	6
	9147	0	TCTTAT CC	GGGTATAGCAGCAGCTGGTA	D6-13	GGGATTTGAGG	15	6

Figure 1. CDR3 regions of V_H6-Cμ cDNAs from 5 young (Y1, Y2, Y3, Y4, Y5) and 5 old (O1, O2, O3, O4, O5) individuals. H, hypermutated clones; NH, non-hypermutated clones. The numbers in the V_H6 nt and J_H nt columns are the number of nucleotides contributed by each gene segment to the CDR3 region. D and J_H gene segments are identified in their respective columns. Potential P nucleotides are underlined in the P/N columns. Bases that do not match identified D segments are indicated in lower case and may represent somatic mutations. Several D segments are listed for clones that matched more than one germline sequence. When a specific D segment could not be assigned, the sequence is presented as consecutive nucleotides in the D column, which includes potential N and D nucleotides.

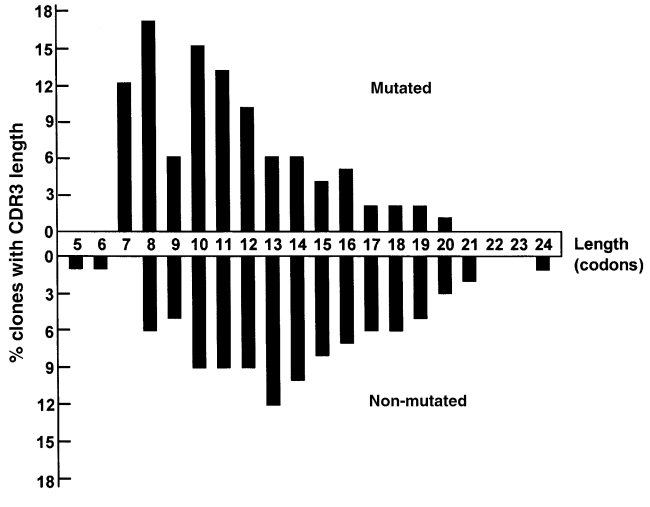

Figure 2. Distribution of amino acid codon lengths in mutated and non-mutated CDR3 sequences.

D gene segment

The D gene segments make the largest contribution to CDR3 length of about 15 nt. The complete sequence of the human D locus by Corbett *et al.*[16] revealed 27 D gene segments that are grouped into seven families. Using LALIGN software at http://www.ch.embnet.org/cgi-bin/LALIGN_form_parser[22] and the criteria described in Materials and Methods, sequences were assigned to D gene segments in 56% (60/108) and 82% (104/127) of the mutated and non-mutated clones, respectively, and in 62% (53/85) and 74% (111/150) of the clones from young and old humans, respectively (Fig. 1). D1–26, D6–13 and D6–19 were the most frequently used segments in all of the V_H6 cDNA libraries (Fig. 3). The most frequently used families in the mutated clones were D6 (38%), D3 (14%), and D1 (20%), and in the non-mutated clones were D6 (43%), D3 (16%), and D1 (16%). There was no significant difference in D usage by either hypermutation status or age. The hydrophilic reading frame of D segments[16] was found in 53% of the 164 heavy chains from mutated and non-mutated clones where a D gene could be confidently assigned, followed by the hydrophobic reading frame in 37% of the clones, and the third frame in 10% of the clones.

The average length of the D gene segments in CDR3 was 12·3 nt in the mutated clones and 14·8 nt in the non-mutated clones ($P=0{\cdot}0006$; Table 1). Differences in nucleotide length between mutated and non-mutated clones by family were as follows: D1, 11·9 versus 12·3; D2, 9·7 versus 17·4; D3, 13·9 versus 17·6; D4, 11·0 versus 15·8, D5, 12 versus 11·6, and D6, 13·0 versus 13·7. Around 10·7 nt were deleted from both the 5′ and 3′ ends of D segments in the mutated clones, and 7·7 nt were deleted in the non-mutated clones. Correlation of D segment length with CDR3 length was significant for the mutated ($P<10^{-4}$) and non-mutated ($P<10^{-4}$) clones. The length of the D segment was not different between clones from young and old humans within the mutated and non-mutated categories.

J_H gene segment

J_H gene segments make a substantial contribution to CDR3 of around 12 nt. There are six functional J_H segments in humans, and all were used in the V_H6-C_μ cDNA sequences shown in Fig. 1. Preferential usage of J_H4 and infrequent usage of J_H1 and J_H2 were found in the cDNA libraries (Fig. 4), as was also observed by others.[9,10,17] A decrease in the use of J_H3 and J_H6 was noted in the mutated (12% and 11%) versus non-mutated (22% and 26%) clones, respectively, and an increase in the use of J_H4 was observed in mutated (62%) versus non-mutated (39%) clones. This difference in J_H usage by mutation status was significant ($P=0{\cdot}001$), but there was no difference by age ($P=0{\cdot}58$). These data contrast with a previous report that

found a difference with age;[23] however, the study analysed DNA clones which partly may have derived from unstimulated B cells. Thus, antigen selection may impose a criterion for expression of certain gene segments.

The average length of the J_H gene segments in CDR3 was 11·1 nt in the mutated clones and 13·9 nt in the non-mutated clones ($P = 0{\cdot}0004$; Table 1). About 4·8 nt were deleted from the ends of the J_H segments in the mutated genes, and 4·5 nt

Table 1. Comparison of nucleotide lengths of CDR3 components

	Mutated		Non-mutated			
Length	Young	Old	Young	Old	Mutated versus non-mutated†	Young versus old‡
Total CDR3§	33.3 (1.5)*	32.7 (1.4)	39.4 (1.8)	41.8 (1.1)	<0.0001	0.55
V_H6 segment§	0.96 (0.11)	0.80 (0.11)	1.15 (0.14)	1.23 (0.08)	0.0023	0.71
D segment¶	12.9 (0.9)	11.7 (0.8)	14.5 (0.9)	15.0 (0.5)	0.0006	0.71
J_H segment§	10.5 (0.8)	11.6 (0.7)	14.0 (1.0)	13.8 (0.6)	0.0004	0.53
P nucleotides¶	0.25 (0.18)	0.25 (0.16)	0.64 (0.2)	0.76 (0.11)	0.0030	0.68
N nucleotides¶						
5′ of D	6.46 (1.15)	6.03 (1.07)	5.88 (1.22)	6.46 (0.68)	>0.9	>0.9
3′ of D	5.00 (0.84)	4.62 (0.79)	5.48 (0.89)	6.10 (0.50)	0.14	0.81

*Standard errors in parentheses.
†P-value for the comparison of mean lengths adjusted for age.
‡P-value for the comparison of mean lengths adjusted for mutation status.
§The analysis included 52 mutated clones from young people, 56 mutated clones from old people, 33 non-mutated clones from young people, and 94 non-mutated clones from old people.
¶The analysis of clones with identified D gene segments: 28 mutated clones from young people, 32 mutated clones from old people, 25 non-mutated clones from young people, and 79 non-mutated clones from old people.

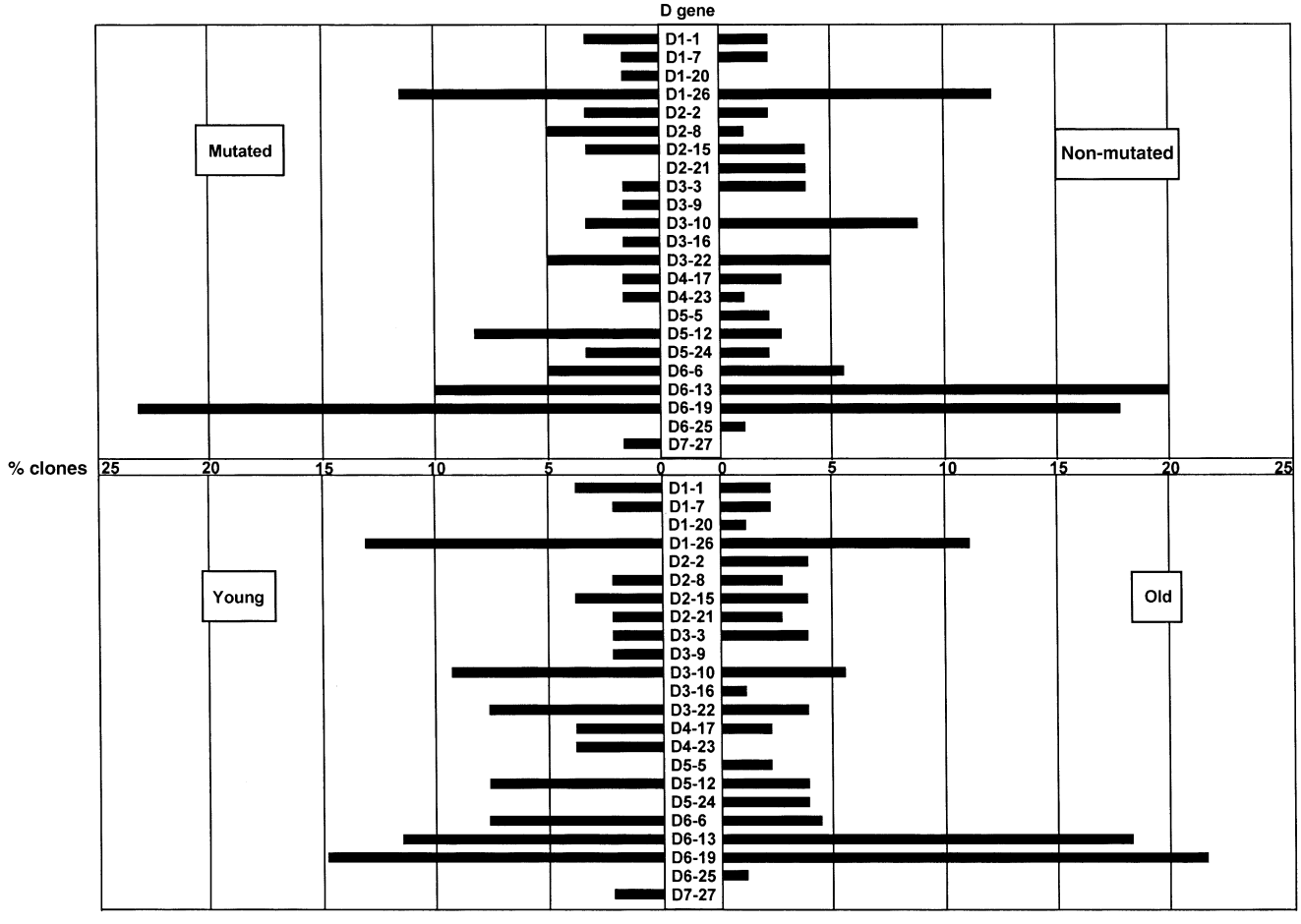

Figure 3. D gene segment utilization in V_H6-Cμ clones. The top half compares usage between all mutated and non-mutated clones, and the bottom half compares usage between all clones from young and old humans.

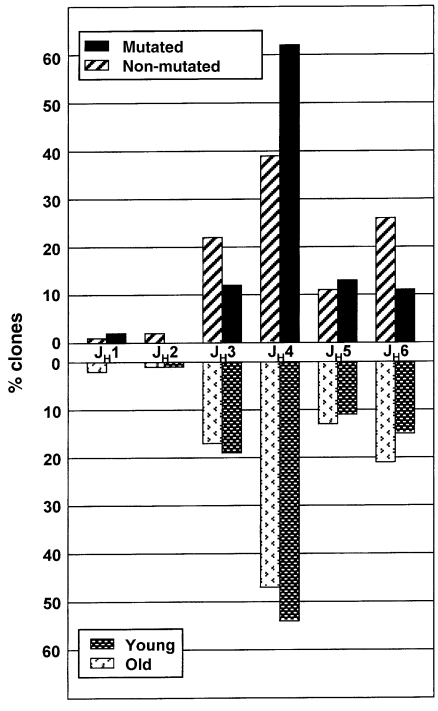

Figure 4. J_H gene segment utilization in V_H6-Cμ clones. The length of the maximum number of germline nucleotides that each J_H segment can contribute to CDR3 is as follows: J_H1, 18 nt; J_H2, 19 nt; J_H3, 14 nt; J_H4, 14 nt; J_H5, 17 nt; and J_H6, 29 nt.

were deleted in the non-mutated genes. Correlation of the J_H length to CDR3 length was also significant in the mutated ($P<10^{-4}$) and non-mutated ($P<10^{-4}$) categories. There was no difference in length between clones from young versus old donors within the mutated and non-mutated groups.

P nucleotides

P nucleotides, which make a minor contribution of 0–4 nt to CDR3, may be present at the ends of V_H, D, or J_H segments in the absence of deletions. Hence, the presence of P nucleotides is dependent on the activity level of exonuclease during VDJ rearrangement. P nucleotides are shown in Fig. 1 and were found flanking the V_H, D, and J_H gene segments. Only 18% of the mutated clones had P nucleotides, compared to 41% of the non-mutated clones. As seen in Table 1, when P nucleotides were present, they were shorter in the mutated clones (0·25 nt) versus non-mutated clones (0·7 nt) ($P=0·003$). The lengths did not differ according to age.

N nucleotides

N nucleotides are inserted by TdT at the V_H-D and D-J_H junctions during joining, and contribute a substantial 12 nt to CDR3 in these clones. N lengths were analysed only in clones with identified D gene segments, and are shown in Fig. 1. As summarized in Table 1, the average length of the N component was similar on both the 5′ and 3′ sides of the D gene, 6·2 nt and 5·3 nt, respectively. The length of N nucleotides was shorter in the mutated genes compared to the non-mutated genes (11·1 nt versus 12 nt). However, since the length located 5′ of D in young individuals was higher in the mutated clones than non-mutated clones, the overall difference was not significant. N nucleotide lengths were similar in genes from the young and old groups within the mutated and non-mutated categories.

DISCUSSION

Mutated VDJ genes have shorter CDR3 lengths than non-mutated genes

Heavy chains with somatic hypermutations have been shown to contain smaller CDR3s than their non-mutated counterparts.[11–13] To identify the components that contribute to the smaller length, we determined the sequence and length of the V_H, D, J_H, P, and N elements in 235 mutated and non-mutated rearranged V_H6 genes from peripheral blood B cells. The overall CDR3 length was decreased considerably in the mutated genes by 8 nt, or around three amino acids, compared to the non-mutated genes. This diminished length could be due to (a) different gene segment usage, and/or (b) varying enzymatic activities of exonuclease and TdT.

Regarding gene segment usage, the V_H6 gene segment was associated with 23 different D gene segments in seven families. Members of the D2 and D3 families are 31 nt long, and the D1, D4, D5 and D6 families are 16–18 nt. However, since there was no difference in D gene usage between the mutated and non-mutated heavy chains, length of the germline D segments was not an explanation for the shorter regions. All six J_H gene segments were used in the V_H6 rearrangements. In the absence of exonuclease, J_H6 can contribute up to 29 nt to CDR3, and the other segments can donate 14–19 nt. As the mutated genes used less of the longer J_H6 segment and more of the shorter J_H4 segment than the non-mutated genes, the shorter CDR3 length is partly due to differential J_H gene usage, which reduced the length by 3 nt or one amino acid. These data confirm those of Brezinschek *et al.*[12] who observed similar results in heavy chains containing predominantly V_H3 gene segments.

Regarding length of the individual components of CDR3, the N, D and J_H parts comprised the bulk of the region by contributing about 10–15 nt each, whereas the V_H and P parts added only 1 nt each. All of these elements were shorter in the mutated genes than non-mutated genes. The following number of nucleotides were deleted by exonuclease in the mutated versus non-mutated genes: V_H segment, 1·1 versus 0·8; D segment, 10·7 versus 7·7; and J_H segment, 4·8 versus 4·5. There were fewer P nucleotides at the ends of V_H, D, and J_H segments in the mutated clones, and when present, their length was shorter than in the non-mutated clones: 0·2 versus 0·7. There were fewer N nucleotides added by TdT in the mutated versus non-mutated genes: 11·1 versus 12. Thus, both exonuclease and TdT enzyme activities contributed to the diminished length of CDR3 in the mutated heavy chains by shortening the five components a total of 5 nt, or almost two amino acids.

Antigen may select for B cells with short CDR3s

Both length and amino acid composition determine the ability of the CDRs to bind antigen with high affinity. Although the lengths of CDRs 1 and 2 are relatively invariant, the length of CDR3 in the heavy chain is extremely diverse. In this data set,

we observed CDR3s ranging from 15 to 72 nt, or 5–24 amino acids. These varied lengths are generated during VDJ joining, and the genes are expressed as immunoglobulin receptors by naive B cells. B cells bearing receptors with short CDR3s may bind to antigen with higher affinity than B cells with receptors with long CDR3s. Selection for cells with short CDR3s is also seen by comparing the length of genes with productive rearrangements (41 nt) to genes with non-productive rearrangements (54 nt) which are not selected.[12,24] In contrast, the length of CDR3 in κ light chains does not vary between mutated and non-mutated clones,[13] perhaps because the length of 6–12 amino acids is optimal for light chains to bind to antigen.

Why would heavy chains with short CDR3s bind antigen better than those with long CDR3s? One intriguing possibility is that since CDR3 is situated in the centre of the antibody combining site, it can fill the cavity with a varying number of amino acids and limit the remaining space for antigen to enter.[25] Thus, antibodies with long CDR3s may have less room in the antibody-binding pocket for antigen to fit. The three-dimensional structures of several antibodies show that long CDR3s fill the antibody-binding cavity and protrude out of it.[26–28] In contrast, antibodies with short CDR3s may have more space in the pocket for antigen to enter and make contact with CDRs 1 and 2 as well. Experimental support for this model is provided by the crystal structures of three anti-lysozyme antibodies complexed with lysozyme.[29–31] As CDR3 shortens, more lysozyme residues come in contact with CDRs 1 and 2.[32] The antigen specificity of the V_H6 heavy chains in this study is not known, but V_H6-encoded antibodies have been shown to bind to bacteria, DNA, and cardiolipin.[33–35]

Once B cells expressing receptors with short CDR3s are selected, the hypermutation mechanism would be activated. B cells bearing receptors with substitutions that change amino acids in the heavy and light chain CDRs 1, 2 and 3 that can bind antigen with even higher affinity are then further selected and expanded. The mutated V_H6 genes in this study were found to have a very high ratio of replacement to silent amino acid changes in CDRs 1 and 2.[15] Thus, there is a correlation between short CDR3s and mutated CDRs 1 and 2, indicating that both have been selected for binding to antigen.

CDR3 length does not change in B cells from adults aged 26–86 years

Fetal B cells from mice and humans have a pauciclonal repertoire of rearranged V gene segments[36,37] and significantly smaller CDR3s, which are primarily due to the limited generation of N nucleotides by TdT.[5–8] The length of the N component in human cells increases with time from around 4 nt at 13 weeks to 15 nt at birth. It is therefore of interest to see if the length of CDR3 and N components changes over many decades of life. Previous studies have reported that heavy chains from young and old people have CDR3s of similar length.[9,10,23] However, these studies included genes that were productively rearranged, non-productively rearranged, mutated, and non-mutated. Since antigen selection for productively rearranged genes with mutations is strongly correlated with diminished CDR3 length, we compared only the productively rearranged genes with mutations to their non-mutated counterparts.

There were fewer mutated clones in the old group, 37%, compared to the young group, 61%, confirming that some aspects of immunity decline with age.[38] Thus, naive B cells with non-mutated antibodies may be generated in old people, but not undergo hypermutation at a high frequency due to impaired T- or B-cell function. In both the mutated and non-mutated categories, there was no difference with age in the length of CDR3 or its individual components. In particular, the length of N nucleotides did not differ in the non-mutated genes, which may arise in newly generated B cells from bone marrow. These results suggest that the expression of TdT in pre-B cells remains constant from the third to ninth decade of life.[39] Thus, fetal B cells may compensate for their restricted V_H usage by expressing immunoglobulin receptors with short CDR3s, which allow CDRs 1 and 2 to come into contact with antigen. Adult naive B cells may express receptors with a wide range of CDR3 lengths to allow the most diversity for making first contact with antigen. Adult memory B cells with mutated receptors may have shorter CDR3s, which allows antigen to interact more effectively with CDRs 1 and 2 in order to initiate hypermutation, which then triggers the logarithmic increases in affinity.

ACKNOWLEDGMENTS

This work was partly supported by a grant from the University of Copenhagen to Karli Rosner, and by the Danish Center for Molecular Gerontology.

REFERENCES

1 Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. Sequences of Proteins of Immunological Interest, 5th edn. Bethesda, MD: US Dept. of Health and Human Services, 1991.
2 Lafaille JJ, DeCloux A, Bonneville M, Takagaki Y, Tonegawa S. Junctional sequences of T cell receptor γδ genes: implications for γδ T cell lineages and for a novel intermediate of V-(D) -J joining. Cell 1989; **59:**859–70.
3 Alt FW, Baltimore D. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc Natl Acad Sci USA 1982; **79:**4118–22.
4 Wood RD, Gearhart PJ, Neuberger MS. Hypermutation in antibody genes. Philos Trans R Soc Lond (Biol) 2001; **356:**1–125.
5 Feeney AJ. Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. J Exp Med 1990; **172:**1377–90.
6 Bangs LA, Sanz IE, Teale JM. Comparison of D, J_H, and junctional diversity in the fetal, adult, and aged B cell repertoires. J Immunol 1991; **146:**1996–2004.
7 Mortari F, Wang J-Y, Schroeder HW Jr. Human cord blood antibody repertoire. J Immunol 1993; **150:**1348–57.
8 Cuisinier AMC, Gauthier L, Boubli L, Fougereau M, Tonnelle C. Mechanisms that generate human immunoglobulin diversity operate from the 8th week of gestation in fetal liver. Eur J Immunol 1993; **23:**110–8.
9 Xue W, Luo S, Adler WH, Schulze DH, Berman JE. Immunoglobulin heavy chain junctional diversity in young and aged humans. Hum Immunol 1997; **57:**80–92.
10 Wang X, Stollar D. Immunoglobulin VH gene expression in human aging. Clin Immunol 1999; **93:**132–42.
11 McHeyzer-Williams MG, McLean MJ, Lalor PA, Nossal GJV. Antigen-driven B cell Differentiation *in vivo*. J Exp Med 1993; **178:**295–307.

12 Brezinschek H-P, Foster SJ, Brezinschek RI, Dörner T, Domiati-Saad R, Lipsky PE. Analysis of the human V_H repertoire. Differential effects of selection and somatic hypermutation on human peripheral $CD5^+/IgM^+$ and $CD5^-/IgM^+$ B cells. J Clin Invest 1997; **99:**2488–501.
13 Brezinschek H-P, Foster SJ, Dörner T, Brezinschek RI, Lipsky PE. Pairing of variable heavy and variable κ chains in individual naive and memory B cells. J Immunol 1998; **160:**4762–7.
14 Berman JE, Mellis SJ, Pollock R *et al.* Content and organization of the human Ig V_H locus: definition of three new V_H families and linkage to the Ig C_H locus. EMBO J 1988; **7:**727–38.
15 Rosner K, Winter DB, Kasmer C, Skovgaard GL, Tarone RE, Bohr VA, Gearhart PJ. Impact of age on hypermutation of immunoglobulin variable genes in humans. J Clin Immunol 2001; **21:**102–115.
16 Corbett SJ, Tomlinson IM, Sonnhammer ELL, Buck D, Winter G. Sequence of the human immunoglobulin diversity D segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, 'Minor' D segments or D-D recombination. J Mol Biol 1997; **270:**587–97.
17 Yamada M, Wasserman R, Reichard BA, Shane S, Caton AJ, Rovera G. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. J Exp Med 1991; **173:**395–407.
18 Sanz I, Kelly P, Williams C, Scholl S, Tucker P, Capra JD. The smaller human VH gene families display remarkably little polymorphism. EMBO J 1989; **8:**3741–8.
19 Demaison C, David D, Letourneur F, Thèze J, Saragosti S, Zouali M. Analysis of human V_H gene repertoire expression in peripheral $CD19^+$ B cells. Immunogenetics 1995; **42:**342–52.
20 Huang C, Stollar BD. Construction of representative immunoglobulin variable region cDNA libraries from human peripheral blood lymphocytes without *in vitro* stimulation. J Immunol Methods 1991; **141:**227–36.
21 van Es JH, Gmelig Meyling FHJ, Logenberg T. High frequency of somatically mutated IgM molecules in the human adult blood B cell repertoire. Eur J Immunol 1992; **22:**2761–4.
22 Huang X, Miller W. A time-effect, linear-space local similarity algorithm. Adv Appl Math 1991; **12:**337–57.
23 van Dijk-Hard I, Soderstrom I, Feld S, Holmberg D, Lundkvist I. Age-related impaired affinity maturation and differential D-JH usage in human VH6-expressing B lymphocytes from healthy individuals. Eur J Immunol 1997; **27:**1381–6.
24 Brezinschek HP, Brezinschek RI, Lipsky PE. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J Immunol 1995; **155:**190–202.
25 Wu TT, Johnson G, Kabat EA. Length distribution of CDRH3 in antibodies. Proteins: Struct Funct Genet 1993; **16:**1–7.
26 Segal DM, Padlan EA, Cohen GH, Rudikoff S, Potter M, Davies DR. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen combining site. Proc Natl Acad Sci USA 1974; **71:**4298–302.
27 Lascombe MG, Alzari PM, Boulot G *et al.* Three-dimensional structure of Rab R19.9, a monoclonal murine antibody specific for the *p*-azobenzene-arsonate group. Proc Natl Acad Sci USA 1989; **86:**607–11.
28 Alzari PM, Spinelli S, Mariuzza RA, Boulot G, Poljak RJ, Jarvis JM, Milstein C. Three-dimensional structure determination of an anti-3-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. EMBO J 1990; **9:**3807–14.
29 Amit AG, Mariuzza RA, Phillips SEV, Poljak RJ. Three-dimensional structures of an antigen-antibody complex at 2.8 Å resolution. Science 1986; **233:**747–53.
30 Sheriff S, Silverton EW, Padlan EA, Cohen GH, Smith-Gill SJ, Finzel BC, Davies DR. Three-dimensional structure of an antibody-antigen complex. Proc Natl Acad Sci USA 1987; **84:**8075–9.
31 Padlan EA, Silverton EW, Sheriff S, Cohen GH, Smith-Gill SJ, Davies DR. Structure of an antibody–antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. Proc Natl Acad Sci USA 1989; **86:**5938–42.
32 Kabat EA, Wu TT. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. J Immunol 1991; **147:**1709–19.
33 Logtenberg T, Young FM, van Es JH, Gmelig-Meyling FH, Alt FW. Autoantibodies encoded by the most J_H-proximal human immunoglobulin heavy chain variable region gene. J Exp Med 1989; **170:**1347–55.
34 Settmacher U, Jahn S, Siegel P, von Baehr R, Hansen A. An anti-lipid A antibody obtained from the human fetal repertoire is encoded by V_H6-$V_\lambda1$ genes. Mol Immunol 1993; **30:**953–4.
35 Andris JS, Brodeur BR, Capra JD. Molecular characterization of human antibodies to bacterial antigens: utilization of the less frequently expressed V_H2 and V_H6 heavy chain variable region gene families. Mol Immunol 1993; **30:**1601–16.
36 Yancopoulos GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, Alt FW. Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. Nature 1984; **311:**727–33.
37 Schroeder HW Jr, Hillson JL, Perlmutter RM. Early restriction of the human antibody repertoire. Science 1987; **238:**791–3.
38 Globerson A, Effros RB. Ageing of lymphocytes and lymphocytes in the aged. Immunol Today 2000; **21:**515–21.
39 Nunez C, Nishimoto N, Gartland GL, Billips LG, Burrows PD, Kubagawa H, Cooper MD. B cells are generated throughout life in humans. J Immunol 1996; **156:**866–72.