Molecular Immunology, Vol. 30, No. 4, pp. 395–402, 1993
Printed in Great Britain.

0161-5890/93 $6.00 + 0.00

JUNCTIONAL DIVERSIFICATION IN THE GENERATION OF THE PRECURSOR OF A DISCRETE IMMUNE RESPONSE

JULIA GEORGE,*† KEVIN M. SHEEHAN,†‡ PETER H. BRODEUR‡ and J. LATHAM CLAFLIN*§

*Department of Microbiology and Immunology, University of Michigan Medical School, 6740 Medical Science II, Ann Arbor, MI 48109-0620, U.S.A.; ‡Department of Pathology, Tufts University School of Medicine, Boston, MA 02111, U.S.A.

(*First received* 20 *March* 1992; *accepted in revised form* 20 *August* 1992)

Abstract—Phosphocholine (PC)-specific antibodies that arise in the mouse in response to *Proteus morganii* (PM) and use V1-DFL16.1-JH1 are characterized by a number of recurring mutations. Most striking is an invariant A for G substitution in codon 95 of VH which results in an asparagine instead of aspartate at that position. Because of the apparent importance of this substitution in an anti-PC(PM) response, we wanted to determine the molecular basis for this base change. A cDNA library derived from pre-immune splenic B cells was examined for the frequency of VDJ containing the A substitution at 95 and the presence of additional point mutations in these sequences. Six different cDNA were isolated which contained an A substitution at the VD junction (frequency 0.00009); a seventh positive cDNA could not be examined. The V segments of four of these cDNA matched known germline genes and were, therefore, unmutated. Two others closely matched V in families whose members have not all been characterized, hence, it is not known whether the mutations observed are somatic or germline in origin. Sequences of 35 cDNA clones, containing the same V segment but differing in D, J and junctional nucleotides, revealed no mutations. These results indicate that the A substitution generated at codon 95 is most likely a product of V–DJ joining.

INTRODUCTION

The murine immune system has the potential to generate greater than 10^{10} unique antibodies as a result of combinatorial diversity (Tonegawa, 1983). Despite this potential the response to some antigens, such as to the phosphocholine (PC) epitope of the Gram negative bacterium *Proteus morganii* (PM), is extremely restricted (Claflin *et al.*, 1989). In total we have sequenced more than 50 antibodies derived from 15 mice (Claflin *et al.*, 1989; Dell *et al.*, 1989; Dell and Claflin, 1990 and unpublished). All of these antibodies use *V*1 of the S107 VH gene family and JH1 in their heavy chains and VK8 and JK5 in their light chains. Moreover, 83% of the antibodies use the DFL16.1 D gene segment. All V–D, D–J, and V–J recombination points are identical or nearly so. Another feature of these antibodies is that they all exhibit recurring somatic point mutations in both CDR2 and CDR3 of the heavy chain. Mutations in CDR2 are always accompanied by a replacement substitution in codon 95, resulting in an asparagine instead of aspartate. This substitution is an identical A for G in the first position of the triplet in 100% of the sequenced antibodies suggesting a need for 95 asparagine in antigen binding or L chain pairing or both. Direct evidence for a functional role of the amino acid at position 95 is the finding that VK8/V1 anti-PC antibodies bind antigens [except the PC(PM) antigen] well if there is an asparagine at 95, but poorly if aspartate is present (Kenny *et al.*, in press). Binding to the PC(PM) antigen requires mutations in CDR2 as well as asparagine at 95 (Claflin *et al.*, 1989).

This last observation results in a biological paradox. Current models describe somatic hypermutation as a mechanism which is not active until after antigenic stimulation (Griffiths *et al.*, 1984; Levy *et al.*, 1989; Wysocki *et al.*, 1986). If this is the case, and the germline encoded specificity is completely unreactive with the antigen, then how is a PC(PM) immune response initiated? The problem might be simplified if we could determine the molecular basis for either of the recurring substitutions, i.e. whether or not both arose as somatic point mutations. The most notable, of course, is the obligate 95 aspartate to asparagine substitution (95N) in CDR3. Because of the proximity of the 95N substitution to the V–D junction, it is logical to think that the substitution is an insertion arising during V–D joining. However, it is not obvious how this can be the case. In the first place, the A for G base change is not the characteristic G or C insertion mediated by TdT (Basu *et al.*, 1983), nor is it a P-nucleotide (Lafaille *et al.*, 1989). Second, junctional sequences in antibodies to other antigens vary, sometimes considerably, even when they are important in antigen binding (Blier and Bothwell, 1987; Cumano and Rajewsky, 1986; Milner *et al.*, 1986; Wysocki *et al.*, 1986; Feeney *et al.*, 1988). In none of these cases are there invariant insertions.

†Current addresses: J. George, Department of Biochemistry, University of Texas, Southwestern Medical Center, Dallas, Texas, U.S.A.; K. Sheehan, Department of Pathology, Stanford University School of Medicine, Palo Alto, California, U.S.A.

§Author to whom correspondence should be addressed.

Abbreviations: PC, phosphocholine; PM, *Proteus morganii*.

To approach the problem we examined the pre-immune repertoire for the frequency of the substitution leading to a G to A substitution in different V–D combinations. The results showed, surprisingly, that this type of substitution was relatively common, occurring at a frequency of about once in 10,000 VDJ. Moreover, it occurs at the V–D junction in the absence of other mutations in V, D or J. This result argues strongly that the single base substitution at codon 95 is an insertion arising during joining of V to D. This, then, defines the germline precursor for anti-PC(PM) antibodies upon which antigen binding somatic variants can be generated by point mutations at one or more positions in CDR2.

MATERIALS AND METHODS

cDNA cloning

A cDNA library representing the adult murine repertoire was generated and extensively characterized elsewhere (Sheehan and Brodeur, 1989). Briefly, the spleens of five adult mice were pooled and stimulated *in vitro* with 50 μg/ml LPS for 3 days. Poly(A)$^+$ RNA was prepared and cDNAs were generated by priming with an oligonucleotide derived from CH2 of the mu heavy chain. Ninety-six *V*1 positive cDNA clones were identified with an oligomer that recognizes codons 44–50 of the BALB/c *V*1 gene: *V*1 5′-TGCAGCAATCCACTCCAGTCT. Hybridization conditions have been described (Geliebter *et al.*, 1986).

Lambdaphage minipreps

Host strain C600 was inoculated with isolated 95N (see Fig. 1) or *V*1 oligomer-positive agar plugs and grown overnight in the presence of 100 mM $MgCl_2$. Bacterial lysates were cleared of cellular debris and treated with DNaseI to remove bacterial DNA. Five microlitres of diethylpyrocarbonate was added to 400 μl of lysate. This was followed by addition of 10 μl 10% SDS and 50 μl of 2.0 M Tris and 0.2 M EDTA (pH 8.5). Lysates were then heated at 68°C for 10 min. Next, 20 μl of 5.0 M KOAc was added. Following 30 min on ice, the precipitate was centrifuged. Supernatants were harvested and extracted once with $CHCl_3$:phenol (1:3). DNA was precipitated and washed three times and resuspended in 200 μl of 10 mM NaCl (Davis *et al.*, 1980).

Polymerase chain reaction

The PCR procedure (adapted from Perkin-Elmer Cetus literature) was carried out in 100 μl reaction volume overlaid with 50 μl of mineral oil. Reaction conditions were as follows: 50 mM KCl; 10 mM Tris, pH 8.3; 2 mM $MgCl_2$; 0.1% gelatin; 2 mM each of dATP, dCTP, dGTP, dTTP; 50 pmoles of each primer; and 1 U of *Taq* polymerase. Reactions were heat-inactivated before addition of enzyme. Each amplification contained a primer from CH1 of the heavy chain, Cmu 5′-GCAGGCGCAGCGGGA, and one of three VH region primers. Oligonucleotides cross-reactive for seven V gene families were used: V123, 5′-GGTGCAGCTGCAGCAGTC; V4, 5′-GGTGAAGCTTCTCGAGTC; and V567, 5′-GAAGCTGGTGGAGTCTGG. These oligomers were selected after examining the 5′ terminal sequences of the seven VH gene families (Brodeur and Riblet, 1984) for the minimum number of higher order consensus sequences. For example, oligomer V123 cross-reacts with the VH1, VH2 and VH3 subgroups, and has no more than two mismatches with any one of them. *V*1 cDNA clones were amplified with the V567 primer.

cDNA clones	VH-D alignment	N	JH
95N	TT TAT TAC TGT GCA AGA AAT		
V1	TT TAT TAC GTG GCA AGA gat gca cacagtgagagg		
14B6	-- --- --- --- --- --- A-- --- --- --- --- --- ---		JH1
FL16.1	gtt tAT TAC TAC GGT AGT AGC TAC C		
V1	TT TAT TAC GTG GCA AGA gat gca cacagtgagagg		
A9	-- --- --- --- --- --- A-- --- --- --- --- --- ---	ctc	JH3
FL16.1	gtt tAT TAC TAC GGT AGT AGC TAC C		
36-60	CA TAT TAC TGT GCA AGA ta cacagtgtgag		
B1	-- --- --- --- --- --- A-- --- --- --- --- --- ---		JH3
FL16.1	gtt tAT TAC TAC GGT AGT AGC TAC C		
V11	CT TAT TAC TGT GCA AGA gat a cacagtgaggg		
C10	-- --- --- --- --- --- AA- --- ---		JH2
FL16.2	att acT ACG GCT ACC		
VH7183	TG TAT TAC TGT GCA AGA ca cacactgagcaaa		
E2	-- --- --- --- --- --- A-- ---	g	JH3
Sp2.2	tac tAT GAT TAC GAC		
VH441	TT TAT TAC TGT GCA AGA cc cacagtgaggaaa		
B10	-- --- --- --- --- --- A-- --- --- ---		JH3
Sp2.2	tac tAT GAT TAC GAC		
VH5	TC TAT TAC TGT GCA AGA cacagtgg		
F1	-- --- --- --- --- --- A-- --- --- -		JH3
Sp2.2	tac tAT GAT TAC GAC		

Fig. 1. Nucleotide sequences of cDNA clones positive with 95N probe are shown compared to both contributing V and D genes. Coding sequences of contributing genes are given in upper case letters. Lower case letters represent flanking sequences. Heptamer is underlined. Within the sequences of cDNA clones, "—" indicates consensus with given upper case, coding sequence. *N*-region nucleotide sequences are given in lower case letters. JH gene usage is indicated.

Each amplification consisted of 30 cycles of the following: denaturation at 95°C for 1 min, hybridization at 50°C for 1.5 min, and extension for 1 min at 72°C. The final extension was continued for 10 min at 72°C.

Sequencing of cDNA clones

DNA from the optimal amplification was excised from 0.8% low melting point agarose (FMC Seaplaque) and sequenced directly using Sequenase. Three to five μl of template DNA in low melting point agarose was used in each 30 μl sequencing reaction (Weber *et al.*, 1991). Sequences were visualized by denaturing in formamide dyes and electrophoresing the bands in a 6% denaturing polyacrylamide gel.

RESULTS

To determine the frequency with which the 95N substitution was expressed in the adult repertoire, a cDNA library representing the pre-immune repertoire in adult mice was generated (Sheehan and Brodeur, 1989). The library was probed with an oligonucleotide, 95N (see Fig. 1) which represents a sequence in the V–D region of 'mutated' PC(PM)-specific antibodies. 200,000 plaques were screened with the 95N probe. Six positives were isolated (Table 1). A *V*1 positive sub-library, consisting of 96 independent isolates, was also screened with the 95N probe and yielded one positively hybridizing clone (A9).

Five of the six 95N positive clones detected in the cDNA library and A9 could be sequenced (Fig. 1). The sequences originate from members of five different V gene families, three different D gene and three different J gene segments. In five of six clones there was a single untemplated A introduced at the V–DJ recombination point. In C10 there are either two As introduced, or none in which case the second A in the AAT codon is a point mutation. No other mutations were seen in D or J of any of these six 95N-positive cDNA clones. Further sequence analysis showed that the V segment encoded portion of four of the six cDNA clones (A9, C10, E1, E10) corresponded to known germline V genes in Genbank. The V of these four cDNAs were 100% homologous to *V*1 (250 bp) (Crews *et al.*, 1981), *V*11 (284 bp) (Crews *et al.*, 1981), 36–60 (255 bp) (Near *et al.*, 1984), and VH441 (268 bp) (Ollo *et al.*, 1981). F1 has four mismatches with the germline VH5 pseudogene (251/255) (Fig.2) (Loh *et al.*, 1983). E2 has two mismatches with a published unproductively rearranged heavy chain gene, BFL1, (243/245) (Fig. 2) (Lawler *et al.*, 1987). Because not all germline members of the gene families to which F1 and E2 belong have been sequenced, we do not know if F1 and E2 represent other germline genes or if they are somatically mutated.

Table 1. Frequency in pre-immune repertoire of VDJ containing AAT at 95

Screened	95N+	Frequency[a] (%)
V1 clones:	1/96	0.01[b]
library:	$6/2 \times 10^5$	0.009[d]
V1-DFL16.1-JH1[c]	$0/2 \times 10^5$	<0.0015

[a]Frequency is calculated with respect to J_H positive clones which is 33.9% of total library (Sheehan and Brodeur, 1989).

[b]V1 genes were determined to be expressed in 1% of total Cmu positive clones examined (Sheehan and Brodeur, 1989), and 1% of V1 clones were positive with 95N probe.

[c]Based on sequence analyis of 95N hybridizing clones. No clones were found which expressed the 95N mutation in the context of a V1-DFL16.1-JH1 H chain.

[d]Our oligonucleotide does not recognize all VH genes, hence the frequency of a single A–T base pair substitution throughout the pre-immune repertoire is an underestimate.

To examine the question of somatic point mutations further, 35 cDNA clones from the *V*1 sub-library were sequenced. These sequences extended from J through CDR2 or a minimum of 210 bp/cDNA, and totaled over 7000 base pairs. All four J region genes were represented (Fig. 3). Extensive junctional diversity was seen. D genes included FL16.1, FL16.2, Q52, Sp2.2, Sp2.5, Sp2.7 and Sp2.8. No non-germline sequences were seen except at the V–D or D–J junctions, indicating that the LPS stimulated B cells from which the cDNA library was generated were not actively mutating (Manser, 1987). Finally, all of the clones sequenced appear to have functional rearrangements in as much as no stop codons are introduced by rearrangement. Of those with a D gene, 87% use reading frame I. Given the size of the data set, this is consistent with the frequency of productive rearrangements seen by Decker *et al.* (1991).

DISCUSSION

The murine immune response to the PC moiety of *P. morganii* has previously been characterized as being highly restricted, even in a primary immune response. In particular, mutations in CDR2 and CDR3 and a specific recurring base change in codon 95 are invariably seen in the heavy chain. Because only a single base is changed relative to the canonical T15 heavy chain (Claflin and Berry, 1988; Gearhart *et al.*, 1981), and because this base change is at the junction between V and D, it is not immediately apparent what the molecular basis for this change is. Antigen-driven selection by *P. morganii* does not appear to act on this substitution, hence we could explain the initiation of an anti-PC(PM) immune response more simply if we could determine the molecular mechanism for this change.

Screening a pre-immune cDNA library did not yield any anti-PC(PM) heavy chains (frequency < 0.0015%), but this can be expected. Decker *et al.* (1991) have estimated that the contribution to diversity of D gene usage and junctional sequences for a single VH–JH pair is greater than 10^4. Nevertheless, we do gain some insight into the generation of the substitution at codon 95 of the heavy chain. Sequence analysis of the 95N positive clones showed no evidence of somatic mutations in V, D

```
cDNA,F1  TACTGGGGCC TCAGTGAAGA TGTCCTGCAA GGCTTCTGGC TACACCTTTA CTAGCTACAC
VH5      ---------- ---------- ---------- ---------- ---------- ----------
         100
         GATGCACTGG GTAAAACAGA GGCCTGGACA GGGTCTGGAA TGGATTGGAT ACATTAATCC
         ---------- ---------- ---T------ ---------- ---------- ----------
         160
         TAGCAGTGGT TATACTAATT ACAATCAGAA GTTCAAGGAC AAGGCCACAT TGACTGCAGA
         ---------- --------C- ---------- ---------- ---------- ----------
         220
         CAAATCCTCC AGCACAGCCT ACATGCAACT GAGCAGCCTG ACATCTGAGG ACTCTGCAGT
         ---------- ---------- ---------- ---------- ---------- ----C--G--
         280
         CTATTACTG TGCAAG
         --------- ------
         340       354

cDNA,E2  GGTCCCGGAA ACTCTCCTGT GCAGCCTCTG GATTCACTTT CAGTAGCTTT GGAATGCACT
BFL1     ---------- ---------- ---------- ---------- ---------- ----------
         35
         GGGTTCGTCA GGCTCCAGAG AAGGGGCTGG AGTGGGTCGC ATACATTAGT AGTGGCAGTA
         ---------- ---------- ---------- ---------- ---------- ----------
         95
         GTACCATCTA CTATGCAGAC ACAGTGAAGG GCCGATTCAC CATCTCCAGA GACAATCCCA
         -----C--C- ---------- ---------- ---------- ---------- ----------
         155
         AGAACACCCT GTTCCTGCAA ATGACCAGTC TAAGGTCTGA GGACACGGCC ATGTATTACT
         ---------- ---------- ---------- ---------- ---------- ----------
         215
         GTGCAAGA
         --------
         275   282
```

Fig. 2. Novel VH gene sequences cloned from a cDNA library with the 95N probe. Sequences are given for cDNA clones and compared to the best matched sequence found in Genbank. Nucleotide mismatches are indicated. Numbers indicate nucleotide positions of known sequences in Genbank. The first amino acid position in VH5 is 13 and in BFL1 is 16 according to Kabat's designation. VH5 represents VH5 pseudogene; BFL1 represents a MOPC21-like gene generated from an Ableson transformed pre-B cell line.

or J. Sequences of an additional 35 *V*1 expressing cDNA clones, which were otherwise as diverse as the 95N positive clones, also revealed no somatic mutations. These data argue strongly that somatic hypermutation is not the molecular basis for the 95N substitution. Because different V and D genes combine to generate 95N-positive clones, we conclude that the mechanism involved is not specific for only *V*1 and its associated D elements. The recombination point of the V genes, varied from 0 to 6 base pairs from the heptamer which is a predicted observation if normal cutting and exonucleolytic nibbling occurred before ligation. We do not see systematic introduction of the A–T base pair for another specific base pair. For these reasons we conclude that the A–T base pair is introduced during V–DJ joining.

The *V*1 gene is part of a small VH gene family consisting of only four members all of which have been cloned from the germline (Crews *et al.*, 1981). Therefore, it did not seem likely that a previously uncharacterized V was responsible for the substitution at codon 95. Nevertheless, we probed genomic DNA with the 95N oligo using appropriately mutated hybridomas as controls. Not surprisingly, we detected no bands (data not shown).

A number of groups have reported a strong bias for the use of reading frame I of DFL and DSP genes in the expressed heavy chain of pre-B cells as well as mature B cells (Decker *et al.*, 1991; Feeney, 1990; Ichihara *et al.*, 1989; Forster *et al.*, 1988). More recent evidence indicates that this bias is a result of selection against reading frames II and III (Gu *et al.*, 1991). All six 95N reactive cDNA clones, as well as all PC(PM) reactive heavy chain genes (DFL16.1+ and DFL16.1−) use the first reading frame of their respective D genes. Introduction of the A–T base pair between V and DJ preserves the use of the first reading frame. Therefore, the 95N "substitution" may actually be an insertion or mutation that occurs during V–DJ joining. Thus, the "germline" heavy chain precursor for an anti-PC(PM) immune response contains a junctional residue which results in the generation of an asparagine at codon 95.

The junctional residue at codon 95 is not a P-nucleotide. Our six cDNA sequences and hybridoma sequences violate the P-nucleotide model on three basic points (Lafaille *et al.*, 1989). We do not see usage of entire D genes. We see various amounts of chewing back from the heptamer in the V genes. Also, the appropriate sequences are not present in the appropriate positions to introduce the A–T base pair between *V*1 and DFL16.1.

Somatic gene conversion could be invoked to explain the generation of the substitution at 95, although there is no evidence that this generates point mutations in murine V(D)J (Chien *et al.*, 1988; Wysocki *et al.*, 1990). Still, one can hypothesize that gene conversion might introduce single base changes in V (Maizels, 1989), so this mechanism remains a formal possibility.

TdT mediated insertion of an *N*-region nucleotide is a possible source for the A–T base pair that appears between *V*1 and DFL16.1. However, TdT is known to have a strong preference for the insertion of Gs and Cs (Basu *et al.*, 1983). While this characteristic of TdT does not eliminate the enzyme as a source for the 95N codon, there are other mechanisms to be considered.

A number of eukaryotic DNA polymerases have been shown to add an untemplated A to the 3′ end of a double stranded oligonucleotide. In the presence of other nucleotide triphosphates, most DNA polymerases will almost exclusively add a dATP (Clark, 1988). If the

(a)

	V1 GCA AGA GAT GCA	N	D	N	JH1 C TAC TGG TAC TTC GAT GTC 101	D-gene usage
B1	--- --- --- --		T AGG		--- --- --- --- ---	SP2.6c,7c,8c
H8	--- --- --- --		T AGG		--- --- --- --- ---	SP2.6c,7c,8c
C10	--- --- --	gag	C TAT GGT AAC TAC	a	-- --- --- --- ---	SP2.5,7
D1	--- --- --	gag	C TAT GGT AAC TAC	a	-- --- --- --- ---	SP2.5,7
B8	--- --- -		AT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
C2	--- --- -		AT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
D7	--- --- -		AT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
F5	--- --- -		AT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
H5	--- --- -		AT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
E10	--- --- -	cc	TAT TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1
B6	--- --- --- -	tt	**TAT GAT TAC G GT** AGC	--- --- --- --- --- ---		**SP2.2/FL16.1**
I7	--- --- --- -	cccc	T TAC TAC GGT AGT AGC	--- --- --- --- --- ---		FL16.1

	V1 GCA AGA GAT GCA	N	D	N	JH2 AC TAC TTT GAC TAC 101	
A2	--- ---			cc	- --- ---	
A3	--- --- -			tg	--- ---	
A6	--- --- --- ---	gg	G TAT GGT A	c	- --- --- ---	SP2.8
C1	--- --- --- --g		ATT ACT ACG GG		- --- --- --- ---	FL16.1
C3	--- --- -	gt	TAT GGT AAC TAC C	c	- --- --- --- ---	SP2.5,7,8

(b)

	V1 GCA AGA GAT GCA	N	D	N	JH3 CC TGG TTT GCT TAC 101	
A5	--- --- -	g	C TAT GGT AAC TAC	gtccc	- --- --- ---	SP2.5,7
A7	--- --- --- -	gg	ACT GGG A		-- --- --- ---	Q52
A9	--- ---	a	AC TAC GGT AGT AGC	ctc	--- --- ---	FL16.1
B2	--- --- ---		CAT TAC	gggc	-- --- --- ---	FL16.2
E2	--- --- ---		CAT TAC	gggc	-- --- --- ---	FL16.2
B3	--- --- -ct		TAC TAT GGT AAC		--- --- ---	SP2.5,7
C5	--- --- -ct		TAC TAT GGT AAC		--- --- ---	SP2.5,7
E4	--- --- ---			cc	- --- --- ---	
I4	--- ---		ccaaggtata		--	?
J1	--- --- --- ---		tgttgtctataggtacgacgg		--- --- --- --- ---	?

	V1 GCA AGA GAT GCA	N	D	N	JH4 AC TAC TAT GCT ATG GAC TAC 101	
A1	--- --- --- ---		ACT ATG ATT AC	ccc	- --- --- --- --- ---	SP2.2
A4	--- --- ---	cccnc	C TAT G	c	- --- --- --- --- ---	SP2.*
B4	--- ---		ACT ATG A	cccc	- --- --- --- --- ---	SP2.2
C4	--- --- --- -	a	A ACT GGG	ct	- --- --- --- --- ---	Q52
D2	--- --- --- --	ccc	C TAC TAT GGT AAC	cc	- --- --- --- --- ---	SP2.5,7
D3	--- --- --- ---	ct	C TAT GGT AAC	ag	- --- --- --- ---	SP2.5,7
E3	--- -		cg ggc cgg gta		--- --- --- ---	?
I6	--- --- ---	c	CT ACG G	cggg	- ---	FL16.1
J4	--- --- --- ---		ACT ATG ATT AC	ccc	- --- --- --- --- ---	SP2.2

Fig. 3. Sequences of V1-positive cDNA clones grouped by JH gene usage. 3′ sequence of *V*1 up to heptamer and 5′ sequence of respective JH element up to heptamer are given. "—" indicates extent of agreement with *V*1 and JH genes. Sequence of D gene is given. In some cases D gene usage could not be completely identified, e.g. SP2.5,7 indicates that D gene could have come from DSP2.5 or DSP2.7. SP2.* indicates that any DSP D element could have been used. The D region for B6 appears to have been derived by D–D joining of Sp2.2 (bold) and FL16.1. N-nucleotides are given in lower case letters. The genetic origin of D regions for I4, J1, and E3 could not be determined, hence they may be N-nucleotide additions and are shown in lower case letters. Possible P-nucleotides are underlined.

unligated V–DJ intermediate has a significant half-life, then promiscuous DNA polymerases may have an opportunity to add an A to the 3′ end of V. Alternatively, any of the enzymatic steps involved in recombination (cutting, nibbling, ligation) may occasionally result in an abasic sight in which the sugar phosphate backbone is intact but the nucleotide base has been lost. To generate our specific recombination product, the abasic sight would have to be on the non-coding strand between *V*1 and DFL16.1 and would effectively be an apyrimidinic site opposite the germline G in codon 95. Apyrimidinic sites are quite stable, having a half life of 100 hr at neutral pH (Duker *et al.*, 1982). Upon subsequent DNA synthesis, dATP could be introduced opposite the abasic sight by the "A-preference" of DNA polymerase (Randall *et al.*, 1987). "A-preference" of DNA polymerase has similarly been invoked to explain the high frequency of single A–T base pairs introduced in non-immune and oncogenic recombination junctions (Roth *et al.*, 1989).

There is no direct evidence for one or another of these models. The models serve to point out that a variety of mechanisms exist within the B cell which can account for the generation of the 95N substitution. All models involve a high degree of precision. Overall, the generation of the 95N codon in V–DJ is an infrequent event and in V1-DFL16.1, must be a rare event. The selective pressure put on this construct by the antigen must be

extreme, which is apparent by the obligate appearance of 95N in PC(PM) reactive antibodies.

Junctional sequences have been shown to be important, but variable, in a number of other antigen systems. In anti-*p*-azophenylarsonate antibodies, junctional sequences at the V–D and D–J junctions can affect affinity by 10-fold (Parhami-Seren *et al.*, 1989). Feeney and co-workers have shown that *N*-region nucleotides are frequently introduced at both the D–J junction and particularly the V–D junction of M511-type anti-PC antibodies. In contrast to the stringent requirement seen in anti-PC(PM) antibodies, the junctions in M511 antibodies code for up to four amino acids and can occur between V and D and/or between D and J (Feeney *et al.*, 1988). Kenny *et al.* (in press) have shown that junctional residues in each of the three different types of anti-PC antibodies determine binding specificity and responsiveness to PC-antigens. Antibodies that arise in response to oxazalone and use VkOx1 do not have a junctional sequence requirement, but do have a structural constraint that D must be three amino acids long (Berek *et al.*, 1985). In addition, the principle mechanism of diversification of the T cell receptor repertoire is *N*-region insertion at V–D, D–D, and D–J junctions (Davis and Bjorkman, 1988).

Our system is unique in that recombination must occur rather precisely between D and J, and then precisely between V and DJ. Small nucleotide overlaps have been hypothesized to be involved in the alignment of D–J and V–DJ junctions, particularly in the absence of *N*-region insertions (Gu *et al.*, 1990). Recombination between DFL16.1 and JH1 is a good example of this, potentially being mediated by the four nucleotide overlap CTAC that is found at the 3′ end of DFL16.1 and the 5′ end of JH1. Recombination and insertion between V and DJ may be mediated by one of a number of mechanisms.

In the V1–JH cDNA sequences there is an over-representation of DFL16.1. Eight out of 12 JH1 positive clones use DFL16.1, all of which have the same D–J junction. Of these eight, five use the same V–D junction. These five cDNA sequences have the same CDR3 structure as seen in the T15 idiotype. This over-representation may be a result of expansion of a B cell clone while still in the mouse, cloning the same cDNA multiple times, or preferential recombination not only between a particular V and DJ combination, but generation of a preferential recombination product. It is not possible to know what light chain was paired with these heavy chains, or to estimate the effect of *in vivo* expansion of the T15 heavy chain in the five mice prior to generating this library. All of the VDJ cDNA inserts should be of similar size and similar base composition. Thus, these cDNAs should have been cloned and expanded with equal efficiency. The over-represented VDJ cDNA is the same recombination product seen in the idiotypically dominant anti-PC antibody, T15 (Quintans, 1989; Feeney, 1991). Our observation that the T15-like, V1-DFL16.1-JH1 recombination product is over-represented in the pre-immune repertoire is in agreement with the observation that T15-like antibodies represent 48–88% of anti-PC precursors (Gearhart *et al.*, 1975). T15 may be over-represented, because it is a favored recombination product (Feeney, 1991). These observations contrast dramatically with the results obtained when amplification is done using primers specific for the four VH36-60 family members and JH2 in which case no duplicates were seen in 225 independent isolates (Decker *et al.*, 1991).

An interesting observation that supports the model of favored recombination comes from a unique set of PC(PM) specific antibodies generated in our laboratory (Dell and Claflin, 1990). These antibodies also use V1-DFL16.1-JH1 as well as VK8-JK5 but are distinct from other anti-PC(PM) antibodies because they use DFL16.1 in reading frame II, and include P-nucleotides at the junction resulting in an asparagine at codon 95 similar to typical anti-PC(PM) antibodies. Strikingly, these antibodies, lack any other mutations in either their heavy or light chains. These antibodies achieve high affinity PC(PM) binding using germline sequences. Yet this type of antibody, has only been seen in one fusion. One interpretation of this data is that the T15 heavy chain is a favored recombination product, and that insertion to create 95N and subsequent CDR2 mutation, occurs at higher frequency than the rearrangement to create the alternative CDR3 structure that can bind PC(PM) without any mutations in CDR2.

In summary, we have shown that the molecular precursor for the PC(PM) immune response requires a precise junctional diversification event. While one might expect this to be an extremely rare event, it appears to occur to a set of V, D and J genes that are recombined preferentially. As stated above, antibodies with only the substitution in CDR3 are unreactive with the PC(PM) antigen. Thus, the remaining paradox is how the PC(PM) immune response is initiated from such a precursor. Recent evidence from a series of B cell transfectants generated in our lab suggests that a cross-reactive antigen must provide the initial stimulatory signal for a PC(PM) immune response anti-(George *et al.*, submitted).

REFERENCES

Basu M., Hegde M. V. and Modak M. J. (1983) Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. *Biochem. biophys. Res. Commun.* **111,** 1105–1112.

Berek C., Griffiths G. M. and Milstein C. (1985) Molecular events during maturation of the immune response to oxazolone. *Nature* **316,** 412–418.

Blier P. R. and Bothwell A. (1987) A limited number of B cell lineages generates the heterogeneity of a secondary immune response. *J. Immun.* **139,** 3996–4006.

Brodeur P. H. and Riblet R. (1984) The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immun.* **14,** 922–930.

Chien N. C., Pollock R. R., Desaymard C and Scharff M. (1988) Point mutations cause the somatic diversification of

IgM and IgG2a antiphosphorylcholine antibodies. *J. exp. Med.* **167,** 954–973.

Claflin J. L., George J., Dell C. and Berry J. (1989) Patterns of mutations and selection in antibodies to the phosphocholine-specific determinant in *Proteus morganii*. *J. Immun.* **143,** 3054–3063.

Claflin J. L. and Rudikoff S. (1977) Uniformity in a clonal repertoire: a case for a germ-line basis of antibody diversity. *Cold Spring. Harbor. Symp. Quant. Biol.* **41,** 725–734.

Clark J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucl. Acids. Res.* **16,** 9677–9686.

Crews S., Griffin J., Huang H., Calame K. and Hood L. (1981) A single VH gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell* **25,** 59–66.

Cumano A. and Rajewsky K. (1986) Clonal recruitment and somatic mutation in the generation of immunological memory to the hapten NP. *Embo J.* **5,** 2459–2468.

Davis M. M. and Bjorkman P. J. (1988) T-cell antigen receptor genes and T-cell recognition [published erratum appears in *Nature* 1988 Oct 20; **335** (6192): 744]. *Nature* **334,** 395–402.

Davis R. W., Thomas M., Cameron J., St. John T. P., Scherer S. and Padgett R. A. (1980) Rapid DNA isolations for enzymatic and hybridization analysis. *Meth. Enzymol.* **65,** 404–411.

Decker D. J., Boyle N. E., Koziol J. A. and Klinman N. R. (1991) The expression of the Ig H chain repertoire in developing bone marrow B lineage cells. *J. Immun.* **146,** 350–361.

Dell C. and Claflin J. L. (1990) Clonal selection and hypermutation in the generation of immunologic memory. In *Somatic Hypermutation in V-regions* (Edited by Steele E. J.), pp. 69–81. CRC Press, Boca Raton, FL.

Dell C. L., Lu Y. and Claflin J. L. (1989) Molecular analysis of clonal stability and longevity in B cell memory. *J. Immun.* **143,** 3364–3370.

Duker N. J., Hart D. M. and Grant C. L. (1982) Stability of the DNA apyrimidinic site. *Mutat. Res.* **103,** 101–106.

Feeney A. J. (1990) Lack of N regions in fetal and neonatal mouse immunoglobulin V–D–J junctional sequences. *J. exp. Med.* **172,** 1377–1390.

Feeney A. (1991) Predominance of the prototypic T15 antiphoshporylcholine junctional sequence in neonatal pre-B cells. *J. Immun.* **147,** 4343–4350.

Feeney A. J., Clarke S. H. and Mosier D. E. (1988) Specific H chain junctional diversity may be required for non-T15 antibodies to bind phosphorylcholine. *J. Immun.* **141,** 1267–1272.

Forster I., Gu H. and Rajewsky K. (1988) Germline antibody V regions as determinants of clonal persistence and malignant growth in the B cell compartment. *Embo. J.* **7,** 3693–4037.

Gearhart P. J., Sigal N. H. and Klinman N. R. (1975) Heterogeneity of the BALB/c antiphosphorylcholine antibody response at the precursor cell level. *J. exp. Med.* **141,** 56–71.

Gearhart P. J., Johnson N. D., Douglas R. and Hood L. (1981) IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature* **291,** 29–34.

Geliebter J., Zeff R. A., Schulze D. H., Pease L. R., Weiss E. H., Mellor A. L., Flavell R. A. and Nathenson S. G. (1986) Interaction between K^b and Q4 gene sequences generates the K^{bm6} mutation. *Molec. cell Biol.* **6,** 645–652.

Griffiths G. M., Berek C., Kaartinen M. and Milstein C. (1984) Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature* **312,** 271–275.

Gu H., Forster I. and Rajewsky K. (1990) Sequence homologies, N sequence insertion and JH gene utilization in VHDJH joining: implications for the joining mechanism and the ontogenetic timing of Lyl B cell and B-CLL progenitor generation. *Embo. J.* **9,** 2133–2140.

Gu H., Kitamura D. and Rajewsky K. (1991) B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound D mu protein and selection of DH element reading frames. *Cell* **65,** 47–54.

Ichihara Y., Hayashida H., Miyazawa S. and Kurosawa Y. (1989) Only DFL16, DSP2, and DQ52 gene families exist in mouse immunoglobulin heavy chain diversity gene loci, of which DFL16 and DSP2 originate from the same primordial DH gene. *Eur. J. Immun.* **19,** 1849–1854.

Lafaille J. J., DeCloux A., Bonneville M., Takagaki Y. and Tonegawa S. (1989) Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V–(D)–J joining. *Cell* **59,** 859–870.

Lawler A. M., Lin P. S. and Gearhart P. J. (1987) Adult B-cell repertoire is, biased toward two heavy-chain variable-region genes that rearrange frequently in fetal pre-B cells. *Proc. natn. Acad. Sci. U.S.A.* **84,** 2454–2458.

Levy N. S., Malipiero U. V., Lebecque S. G. and Gearhart P. J. (1989) Early onset of somatic mutation in immunoglobulin VH genes during the primary immune response. *J. exp. Med.* **169,** 2007–2019.

Loh D. Y., Bothwell A. L., White-Scharf M. E., Imanishi-Kari T. and Baltimore D. (1983) Molecular basis of a mouse strain-specific anti-hapten response. *Cell* **33,** 85–93.

Maizels N. (1989) Might gene conversion be the mechanism of somatic hypermutation of mammalian immunoglobulin genes? *Trends Genet.* **5,** 4–8.

Manser T. (1987) Mitogen-driven B cell proliferation and differentiation are not accompanied by hypermutation of immunoglobulin variable region genes. *J. Immun.* **139,** 234–238.

Milner E. C., Meek K. D., Rathbun G., Tucker P. and Capra J. D. (1986) Are anti-arsonate antibody N-segments selected at both the protein and the DNA level? *Immun. Today* **7,** 36–40.

Near R. I., Juszczack E. C., Huang S. Y., Sicari S. A., Margolies M. N. and Gefter M. (1984) Expression and rearrangement of homologous immunoglobulin VH genes in two mouse strains. *Proc. natn. Acad. Sci. U.S.A.* **81,** 2167–2171.

Ollo R., Auffray C., Sikorav J. L. and Rougeon F. (1981) Mouse heavy chain variable regions: nucleotide sequence of a germ-line VH gene segment. *Nucl. Acids. Res.* **9,** 4099–4109.

Parhami-Seren B., Wysocki L. and Margolies M. N. (1989) The amino acid residues at the VH–D–JH junctions affect the affinity of anti-*p*-azophenylarsonate antibodies. *J. Immun.* **143,** 4090–4097.

Quintans J. (1989) Cellular competition and the promotion of T15 to idiotypic dominance. *Immun. Rev.* **110,** 119–134.

Randall S. K., Eritja R., Kaplan B. E., Petruska J. and Goodman M. F. (1987) Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. biol. Chem.* **262,** 6864–6890.

Roth D. B., Chang X. B. and Wilson J. H. (1989) Comparison of filler DNA at immune, non-immune and oncogenic

rearrangements suggests mutiple mechanisms of formation. *Molec. cell Biol.* **9,** 3049–3057.

Sheehan K. M. and Brodeur P. H. (1989) Molecular cloning of the primary IgH repertoire: a quantitative analysis of VH gene usage in adult mice. *Embo. J.* **8,** 2313–2320.

Tonegawa S. (1983) Somatic generation of antibody diversity. *Nature* **302,** 575–581.

Weber J. S., Berry J., Litwin S. and Claflin J. L. (1991) Somatic hypermutation of the JC intron is markedly reduced in unrearranged K and H alleles and is unevenly distributed in rearranged alleles. *J. Immun.* **146,** 3218–3226.

Wysocki L., Gefter M. L. and Margolies M. N. (1990) Parallel evolution of antibody variable regions by somatic processes: consecutive shared somatic alterations in VH genes expressed by independently generated hybridomas apparently acquired by point mutation and selection rather than by gene conversion. *J. exp. Med.* **172,** 315–323.

Wysocki L., Manser T. and Gefter M. L. (1986) Somatic evolution of variable region structures during an immune response. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1847–1851.