POLYMORPHISMS IN ANTI-PHOSPHOCHOLINE ANTIBODIES REFLECTING EVOLUTION OF IMMUNOGLOBULIN FAMILIES

By STEPHEN H. CLARKE, J. LATHAM CLAFLIN, MICHAEL POTTER, and STUART RUDIKOFF*

From the Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; and the Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109

The BALB/c immune response to the haptenic determinant phosphocholine (PC)[1] has been extensively characterized at both the protein and nucleic acid levels. All PC-binding myeloma (1–3) and hybridoma (4–6) proteins from conventionally immunized mice express one of three light (L) chains designated T15, M167, and M603. Recent studies comparing the protein sequences of the V_H segments (amino acids 1 $\approx$ 95) from a number of BALB/c anti-PC hybridomas and myelomas (6) to the DNA sequences of genes encoding the anti-PC V_H segments (7, 8) have concluded that the entire V_H repertoire seen in this response is possibly generated from a single V_H gene (designated VI or VT15) and that variations in sequence such as those seen in the M167 and M603 myeloma heavy (H) chains are the result of somatic mutations. This particular V_H gene directly encodes the T15 V_H segment, defining the T15 V_H protein sequence as the BALB/c germline gene product. We have recently reported that, in fact, a second member of the PC-V_H gene family is probably also used in the production of antibodies with this specificity (9). The H chains of BALB/c PC-binding monoclonal antibodies derived from the T15 V_H gene have been shown to use different "D" regions that may vary in primary sequence as well as in length. All of the BALB/c proteins use the J_H1 joining segment (6, 10–12).

We have previously described the amino acid sequence of the entire variable (V) region (amino acids 1–118) from a PC-binding myeloma protein, CBBPC-3 (C3),[2] carrying the C57BL allotype (12). The C3 V_H segment was found to differ at four positions from the T15 V_H segment of BALB/c origin. Furthermore, partial sequences of the V_H segments from antibodies raised in C57BL mice (13) demonstrated the same characteristic substitutions found in C3, which suggests that C3 is the product of a gene in C57BL allelic to T15 of BALB/c. Serologic analysis using anti-idiotypic antisera to T15 and C3 further demonstrated that the two idiotypes segregated as simple Mendelian alleles linked to the immunoglobulin constant region (IgC_H) locus

* To whom correspondence should be addressed at the Laboratory of Cell Biology, Bldg. 8, Rm. 204, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205.

[1] *Abbreviations used in this paper:* CFA, complete Freund's adjuvant; IEF, isoelectric focusing; PC, phosphocholine; PC-KLH, PC-conjugated keyhole limpet hemocyanin; RIA, radioimmunoassay; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

[2] The C3 myeloma protein arose in the congenic inbred CBB-22 strain developed by backcrossing the B (black), C (color), and IgC_H loci from C57BL/Ka onto a BALB/c background for 22 consecutive generations.

(14), which supports the previous conclusions as to the allelic nature of these V_H segments.

The extensive antigenic (allotypic) differences found among six of the immunoglobulin IgC_H regions of BALB/c and C57BL (15), in addition to the V_H differences described above, reflect the considerable evolutionary divergence in the immunoglobulin genes of these two strains. In fact, the genotypes of BALB/c and C57BL are probably as different as any two inbred mouse strains. It is therefore of interest to examine anti-PC monoclonal antibodies of C57BL origin to ask: (*a*) What is the structure of the V region segments used in anti-PC antibodies of C57BL as compared with BALB/c? (*b*) How do these structures reflect evolutionary changes in this family? (*c*) Are the gene segments used in these two divergent genotypes assembled similarly? For this purpose we have prepared and determined the amino acid sequence of PC-binding monoclonal antibodies generated in C57BL mice. In addition, we have also performed analyses of CBA/J hybridomas exhibiting similar specificities. The CBA strain is of particular evolutionary interest in that it is closely related to BALB/c at the IgC_H locus (15), but mice of this strain serologically express the C3 (C57BL) type V_H segment (13, 14).

Materials and Methods

Mice. CB-20, C57BL/6N, and (BALB/c × C57BL/6N)F_1 mice were obtained from National Cancer Institute contract 1-CB2-5548 with Litton Bionetics, Kensington, MD or bred in this laboratory from stock obtained from Litton Bionetics.

Immunizations. Heat-killed *Streptococcus pneumoniae* (R36A) was kindly provided by Rose Lieberman (National Institute of Allergy and Infectious Diseases, National Institutes of Health). Heat-killed *Proteus morganii* and PC-conjugated keyhole limpet hemocyanin (PC-KLH) were the gift of Dr. James Kenny (Department of Microbiology, Uniformed Services University of Health Sciences, Bethesda, MD). Hybridoma 914 was produced from the fusion of spleen cells of two CB20 mice immunized intraperitoneally 4 d earlier with 10^8 heat-killed *S. pneumoniae* in 0.1 ml saline. Hybridoma 1613 was produced from the fusion of spleen cells of two C57BL/6N mice also immunized with *S. pneumoniae*. These mice were immunized intraperitoneally with 2 × 10^8 organisms in 0.2 ml saline followed by 1 × 10^8 organisms in 0.1 ml saline 4 d later. 3 d after the last injection, their spleen cells were combined for fusion. Hybridomas 23169 and 2312 were produced from the fusion of spleen cells of two C57BL/6N mice immunized intraperitoneally with 0.1 ml of *P. morganii* (6.0 optical density/ml) mixed 1:1 with complete Freund's adjuvant (CFA). After 3 wk, the mice were again given 0.1 ml of *P. morganii* in CFA, and 4 d later their spleen cells were combined for fusion. Hybridomas 2851 and 2857 were produced from 2 C57BL/6N mice immunized intraperitoneally with 100 μg of PC-KLH in 0.2 ml of CFA. After 10 d, the mice were given a second injection of 100 μg of PC-KLH in CFA, and 4 d later their spleen cells were combined for fusion. Hybridoma 293 was produced from the fusion of spleen cells of two C57BL/6N mice immunized with PC-KLH in the same manner.

Hybridoma Production. CBA/J hybridomas 140.1C2(1C2) and 101.3C2(3C2) were prepared as described by Claflin et al. (5). All C57BL/6N and CB20 hybridomas were prepared by fusion with the nonsecreting cell line, Sp2/0-Ag14 (16), kindly provided by Dr. George Kohler (Basel Institute for Immunology, Basel, Switzerland). The Sp2/0 cells were maintained in culture and the fusions were performed at a ratio of 1:3–1:10 (Sp2/0: spleen cells), as described by Pawlita et al. (17). Approximately 2 wk after fusion, PC-binding hybridomas were identified by testing the tissue culture supernatants for PC-binding antibody in a solid-phase radioimmunoassay (RIA) using PC-conjugated bovine serum albumin. Cells from positive cultures were cloned once or twice and then 1 × 10^5 cells were injected intraperitoneally into pristane-primed (C57BL/6N × BALB/c)F_1 mice.

H and L Chain Preparation. Ascites containing hybridoma proteins were mildly reduced with 5 mM dithiothreitol for 75 min and alkylated with 8.3 mM iodoacetamide for 20 min. After overnight dialysis against borate-buffered saline, pH 8.0, hybridoma proteins were affinity-

purified on a PC-Sepharose column (18). Proteins were partially reduced and alkylated (19), dialyzed into 6 M urea–1 M acetic acid, and H and L chains were separated by gel filtration on Sephadex G-100 columns equilibrated in 6 M urea–1 M acetic acid.

Cyanogen Bromide Cleavage. H and L chains were cleaved with CNBr at a 4:1 weight ratio (CNBr:protein) in 70% formic acid (10), and the fragments were separated on a Sephadex G-100 column equilibrated in 5 M guanidine–0.2 M NH_4HCO_3. Pooled fractions were dialyzed extensively against 0.2 M NH_4HCO_3 and lyophilized. To separate fragments linked by intrachain disulfide bonds, a complete reduction and alkylation was performed (11). Resulting peptides were separated by gel filtration on a Bio-Rad A 0.5 M agarose column (Bio-Rad Laboratories, Richmond, CA) equilibrated in 6 M guanidine–Tris buffer, pH 8.0. Pooled fractions were dialyzed extensively against 0.2 M NH_4HCO_3 and lyophilized.

Sequence Determinations. Automated Edman degradations were performed on intact chains and CNBr fragments using a modified Beckman 890 C sequencer (Beckman Instruments, Inc., Fullerton, CA) (20, 21) using a 0.25 M quadrol buffer program. Identification of phenylthiohydantoin amino acids was performed by high-pressure liquid chromatography (22).

H Chain Class Determination. The H chain class of the C57BL/6N and CB20 hybridoma proteins was determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in slab gels using purified H chain or whole antibody (23). In addition, the H chain class was verified by Ouchterlony analysis using H chain-specific rabbit antisera obtained from Litton Bionetics.

Determination of Association Constants (K_a). The association constants of hybridoma and myeloma proteins for PC were determined by fluorescence titrations using the procedure of Jolley and Glaudemans (24). Titrations were carried out at 25°C, using an MPF-3L fluorescence spectrometer (Perkin-Elmer Corp., Eden Prairie, MN). The excitation and fluorescence wavelengths were 280 and 330 nm, respectively.

Results

Hybridomas. To assess structural differences occurring in antibodies used in the response to PC in diverse inbred mouse strains, a series of hybridoma proteins was generated and characterized as presented in Table I. Six C57BL/6N hybridomas and one CB20 (a BALB/c congenic strain with the C57BL IgC_H locus) hybridoma were produced from five separate fusions of a non-immunoglobulin-secreting cell line, Sp2/0-Ag14 (16), with spleen cells of PC-immunized mice using the method of Kohler and Milstein (25). An additional series of five CBA/J hybridomas derived from three separate fusions (5, 26) was also examined. The association constant (K_a) of these proteins for PC hapten was determined by fluorescence titration (24) and the range of values obtained is similar to that exhibited by other BALB/c PC-binding myeloma and hybridoma proteins (6, 18, 27). Hybridomas 2857 of C57BL/6N origin and 3C2 and 101.6G6 (6G6) of CBA/J origin did not change fluorescence upon binding PC, and therefore their affinity for PC could not be determined by this method. Although these three proteins did not change fluorescence upon reaction with PC, 3C2 and 6G6 have been shown to bind both PC and a number of PC analogues in an RIA (26) and 2857 was hapten-eluted from a PC immunoabsorbent column, which indicates specificity for the PC determinant. All C57BL/6N proteins characterized in this study are of the IgM class as determined by SDS-PAGE and by Ouchterlony analysis using H chain constant region-specific rabbit antisera. The CBA/J hybridoma proteins are also of the IgM class with the exception of 1C2, an IgG2 protein (5).

L Chain V Regions. L chain amino terminal sequences of the 12 C57BL/6N and CBA/J hybridoma proteins presented in Fig. 1 were obtained by automated degradation of intact chains. All hybridoma L chains fall into three groups based on sequence similarities to the L chains of the BALB/c myeloma proteins T15, M603,

TABLE I
Characteristics of C57BL/6N and CBA/J Hybridoma Proteins

Protein	Strain of origin	Ig class	Immunogen	K_a ($\times 10^5$) M^{-1} for PC
914	CB-20*	IgM	*S. pneumoniae* (R36A)	—‡
1613	C57BL/6N	IgM	*S. pneumoniae* (R36A)	1.8 ± 0.055
23169	C57BL/6N	IgM	*P. morganii*	3.0 ± 0.20
2312	C57BL/6N	IgM	*P. morganii*	0.61 ± 0.043
2857	C57BL/6N	IgM	PC-KLH	NC§
293	C57BL/6N	IgM	PC-KLH	6.3 ± 0.91
2851	C57BL/6N	IgM	PC-KLH	—
3C2	CBA/J	IgM	*S. pneumoniae* (R36A)	NC
6G6	CBA/J	IgM	*S. pneumoniae* (R36A)	NC
6F9	CBA/J	IgM	*S. pneumoniae* (R36A)	6.1 ± 0.30
1C2	CBA/J	IgG2	PC-KLH; PC-BGG	—
7C6	CBA/J	IgM	PC-KLH; PC-BGG	12.0 ± 1.1
C3	CBB-22‖	IgA	—	5.8 ± 0.82
T15	BALB/c	IgA	—	2.3 ± 0.5¶
M603	BALB/c	IgA	—	1.6 ± 0.4
M167	BALB/c	IgA	—	1.2 ± 0.2

* The immunoglobulin congenic strain CB-20 was established by introgressively backcrossing the C57BL IgC_H complex locus into BALB/c for 20 consecutive generations. Backcross-20 mice were mated and mice homozygous for the C57BL constant region determinants were used as stock for the CB-20 strain.

‡ Proteins 914, 2851, and 1C2 were not tested for their affinity for PC.

§ Proteins 2857, 3C2, and 6G6 did not change fluorescence upon binding PC and therefore affinity constants were not determined.

‖ The CBB-22 strain was established by introgressively backcrossing the C57BL IgC_H, C (color), and B (black) loci into BALB/c for 22 generations. Backcross-22 mice were then mated and mice homozygous at these loci were used as stock for the CBB-22 strain.

¶ The affinities for the BALB/c myeloma proteins T15, M603, and M167 are from Chesbro and Metzger (18), and Metzger et al. (27).

and M167 (1, 2). The first group (Fig. 1) consists of proteins 914 of CB20 origin, 293 and 2851 of C57BL/6N origin, and 140.7C6 (7C6) and 1C2 of CBA/J origin. The L chains of these proteins, with the exception of the two C57BL/6N proteins, are identical for the first 50 amino acids to the T15 L chain of BALB/c origin. Proteins 293 and 2851 have a substitution of leucine for methionine at position 4, which can be generated by either a one- or two-base change at the DNA level.

The second group consists of proteins 1613, 23169, and 2312 of C57BL/6N origin and 100.6F9 (6F9) of CBA/J origin. The first 50 positions of the L chains from these proteins are identical to the germline amino acid sequence of the M603-type L chain of BALB/c mice (4, 6). The M603 myeloma L chain appears to have accumulated mutations and differs from the germline amino acid sequence at positions 18 and 32.

The third group includes proteins 2857 of C57BL/6N origin and 6G6 and 3C2 of CBA/J origin. The L chains of these proteins are identical for the first 50 positions to the germline-encoded amino acid sequence of the M167 L chain of BALB/c mice (29).

In addition to the amino terminal sequences, the entire L chain variable region from 2857, an M167-like protein of C57BL/6N origin, was determined (Fig. 2) by degradation of intact L chain and the CNBr fragment containing amino acids 52-175. CNBr-cleaved L chains were chromatographed on a Bio-Rad A 0.5 M agarose

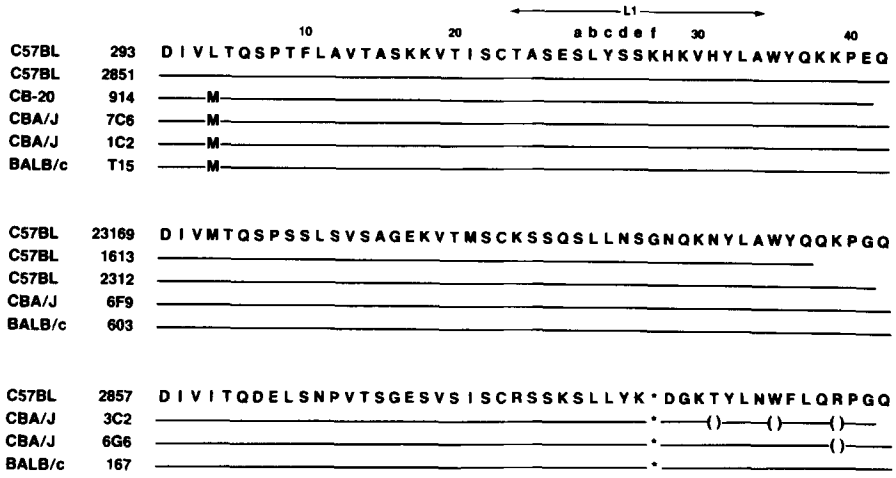

FIG. 1. A comparison of the L chain V region amino terminal sequences from the C57BL/6N and CBA/J hybridoma proteins with the T15, M603, and M167 L chains of BALB/c origin. The M603 L chain sequence is the apparent germline sequence obtained by Kocher et al. (4) and Gearhart et al. (6). The M167 L chain sequence is the amino acid sequence deduced from the germline nucleotide sequence (29). Open parentheses indicate an unassigned amino acid. Asterisk (*) indicates a one amino acid deletion in the M167 group when compared with T15 amd M603. Numbering and hypervariable region assignments in all figures are according to Kabat et al. (28).

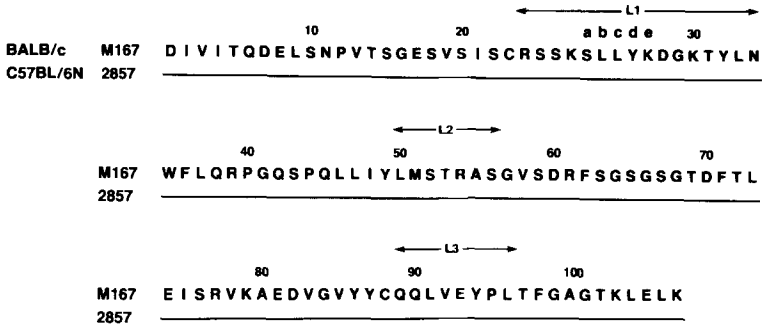

FIG. 2. A comparison of the complete variable region sequence of the 2857 L chain with the M167 germline V_K and J_K five genes of BALB/c (29, 30).

column equilibrated in 6 M guanidine–Tris pH 8.0 (Fig. 3). The fragment containing amino acids 52-175 was identified in peak P1C and was sequenced from position 52-107.

H Chain V Regions. Complete V region sequences were determined for the C57BL IgM PC-binding monoclonal antibodies 293, 1613, 23169, and 2857, as well as amino terminal sequences for proteins 2312 and 914 (Fig. 4). Partial sequences were obtained for the CBA/J hybridomas 1C2 and 3C2 and are presented along with three recently reported (9) complete V region sequences from proteins 6F9, 7C6, and 6G6 (Fig. 5). V region sequences were determined by automated degradation of intact H chains and appropriate CNBr fragments as described by Clarke et al. (9). The V_H segments from the C57BL/6N and CB20 hybridoma proteins, with the exceptions of positions 82a and 95, are identical to the amino acid sequence of the C3 V_H segment (Fig. 4). Position 82a was not conclusively identified in proteins 1613 and 23169 and is

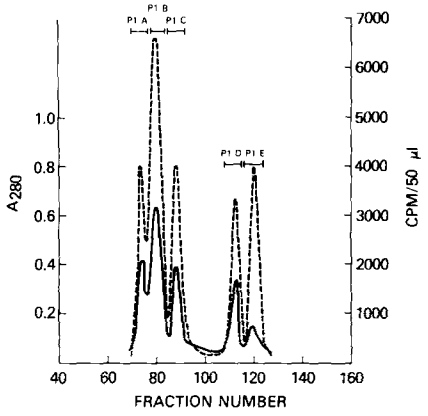

FIG. 3. Sephadex G-100 chromatography of CNBr-cleaved 2857 L chain. Columns were equilibrated in 5 M guanidine 0.2 M ammonium bicarbonate.

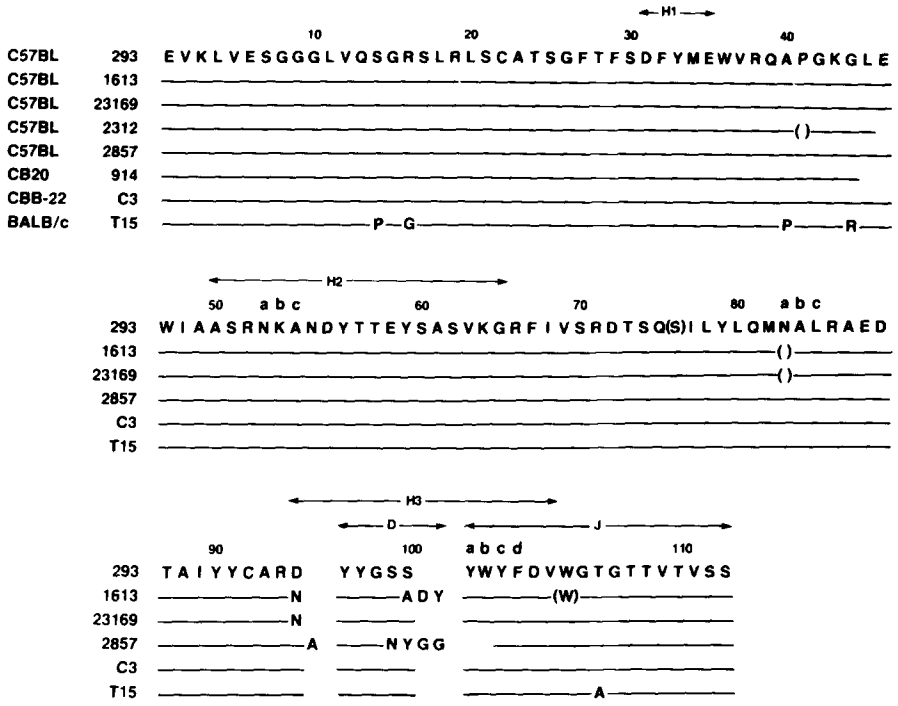

FIG. 4. A comparison of H chain V region sequences from the C57BL/6N hybridoma proteins with the C3 and T15 H chain sequences. Open parentheses indicate an unassigned amino acid. Amino acids in parenthesis were tentatively identified.

presently unassigned. The only position at which the C57BL/6N V_H segments clearly differ is residue 95, located at the V-D junction. Position 95 was observed to be asparagine in 1613 and 23169, whereas aspartic acid was found in 2857 and 293. Amino terminal sequences of six H chains from CBA/J PC-binding hybridoma proteins and the complete H chain V region sequences of three of these (Fig. 5) reveal that proteins 6F9 and 7C6 are identical to the C3 V_H segment with the exception of

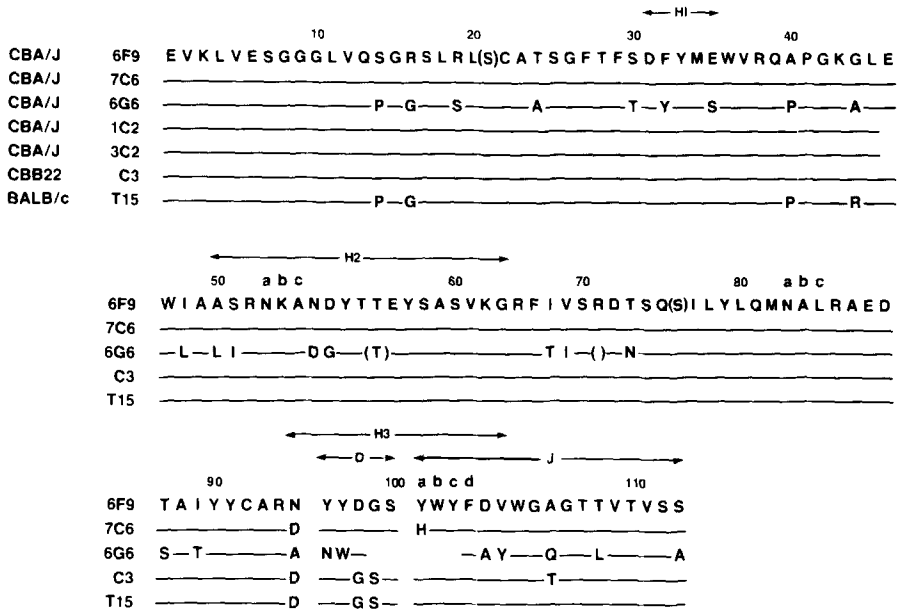

FIG. 5. A comparison of H chain V region sequences from the CBA/J hybridoma proteins with the C3 and the T15 H chain sequences. Amino acids in parentheses indicate a tentative assignment and open parentheses indicate an unidentified position. Serine at positions 21 and 76 was tentatively identified in 6F9 and conclusively identified in all other H chains.

an Asp-Asn interchange at position 95 as seen in both BALB/c and C57BL anti-PC V_H segments. These H chains are therefore the product of a V_H gene that encodes a V_H segment that is structurally identical to that of C57BL/6N mice.

D Regions. Analysis of the complete V region sequences (Fig. 4) reveals that three different D regions are present in the H chains of four C57BL/6N hybridoma proteins. The same D region is found in 23169, 293, and C3, and is identical to that of the T15 H chain as well as six other PC-binding hybridoma and myeloma proteins of BALB/c origin (6, 11).

Two different D regions were found among the CBA/J hybridoma proteins (Fig. 5). H chains of proteins 7C6 and 6F9 each express the C3 V_H segment but have identical D regions that differ from that of C3. Aside from the D region of the T15 H chain, this is the only D region to appear in more than one H chain of PC-binding antibodies.

J Regions. The four H chains of C57BL/6N hybridoma proteins for which complete V region amino acid sequences were determined express a J region similar to the J_H1 of BALB/c mice (Fig. 4). However, in all cases, this J_H region has a substitution of threonine for alanine at position 105, which requires a single-base pair change when compared with the J_H1 sequence of BALB/c mice. The J regions of the CBA/J hybridoma proteins 7C6 and 6F9 are identical to the J_H1 regions expressed in BALB/c proteins (Fig. 5). The J_H region of protein 6G6 is identical to the J_H3 region of BALB/c mice and is the only example of a PC-binding antibody using a J_H region other than J_H1 (Fig. 5).

Discussion

A comparison of the amino acid sequences of PC-binding hybridoma proteins of C57BL/6N, CBA/J, and BALB/c origin reveals several interesting observations relating to the evolution of gene families that encode the humoral response to PC.

V_K and V_H Segments Are Remarkably Conserved. The V_K amino acid sequences determined from these three strains span ~200 amino acids from three distinct V_K subgroups represented by the myeloma L chains T15 (V_{K22}), M603 (V_{K8}), and M167 (V_{K24}) (Figs. 1 and 2). In the T15 subgroup, all L chains are identical for their first 50 amino acids (Fig. 1), with the exception of the two C57BL proteins that have the identical Leu-Met interchange at position 4. Because this substitution appears in both L chains, it is likely to be germline encoded and possibly represents an allelic form of the BALB/c T15 V_L gene. In the M603 subgroup, BALB/c, C57BL, and CBA mice express L chains identical for their first 50 amino acids. In the M167 subgroup, all L chains are identical to the BALB/c germline sequence (29).

In addition to the amino terminal sequences, the entire L chain V region from 2857, an M167-like protein of C57BL origin, was determined (Fig. 2). This V_K region is identical to the amino acid sequence deduced from the germline M167 V_K and J_K genes of BALB/c (29, 30). Therefore, the germlines of BALB/c and C57BL can encode identical M167-like L chains.

Even though "silent" mutations, which would be undetected at the protein level, may have occurred in these three strains, the data nonetheless demonstrate a remarkable conservation of these V_K sequences in the diverse germlines. It has previously been shown (31–34) that the serum response of most inbred strains to PC antigens consists almost entirely of L chains, which, by isoelectric focusing (IEF) analysis, comigrate with the T15, M603, or M167 L chains. The hybridoma sequences obtained in the present study completely support the predicted conservation observed by IEF and indicate that this conservation extends to identical amino acid sequences which in turn reflect the preservation of their respective gene structures.

Excluding the CBA/J 6G6 sequence, the PC-V_H segments from mice of these strains differ by five amino acid substitutions (Figs. 4 and 5) at positions 14, 16, 40, 44, and 95. Position 95 lies at the V_H-D_H juncture and substitutions in this region probably result from alterations in the frame of recombination between V_H and D_H gene segments. Both C57BL and CBA anti-PC monoclonal antibodies use V_H segments that express the identical substitutions that distinguish these strains from BALB/c. These sequences thus define, at the amino acid sequence level, allelic forms of the T15 V_H gene. This observation is consistent with previously reported serological and amino acid sequence data that indicate that CBA/J and C57BL/6N mice express PC-binding antibody with H chains of the same allelic form (13, 14). All of the V_H segment substitutions are located in framework regions, which suggests a conservation of hypervariable region sequences. With the exception of the 6G6 V_H segment of CBA/J origin and probably the HPCG15 V_H segment of BALB/c origin (6), all PC-binding antibodies from mice of these three strains are composed of the same three V_K regions pairing with a single V_H segment. Thus, the same combinatorial association observed in BALB/c antibodies is found to occur in other genotypes. The 6G6 V_H segment, which appears to be the product of a second PC V_H gene (9), combines with an M167 V_K region in CBA/J, and a similar V_H (HPCG15) of BALB/c origin (6) pairs with an M603 V_K region, which indicates that other members of the PC V_H

family can probably also pair with any of these three V_K regions to produce anti-PC antibodies. We have previously reported that the 6G6 V_H (CBA/J origin) amino acid sequence differs from the T15 and C3 V_H sequences at 17 and 20 positions, respectively (Fig. 5). A comparison of this sequence with the amino acid sequences translated from the BALB/c PC-V_H gene family indicated that the 6G6 V_H was not the product of the T15 V_H gene (or its CBA homologue) but derived from a second PC-V_H gene, which, we have hypothesized, may have undergone gene conversion (see Clarke et al. [9] for a detailed discussion).

D Region Diversity. The D segment of the H chain comprises a portion of the third hypervariable region and varies in length from apparently one to several amino acids (7, 35). Because this segment falls within the third hypervariable region, it has been proposed to potentially introduce significant functional antibody diversity by altering binding specificity. In the anti-PC antibodies, multiple D regions are used by mice of each strain. Among the seven complete C57BL and CBA/J V region sequences, five different D segments varying in length and/or sequence (Table II) were observed that apparently represent the products of multiple D genes. Some of this variation can be

TABLE II

*D Region Sequences in PC-binding Antibodies**

Strain	Origin	Sequence	
		V_H	D_H
BALB/c	Germline DFL 16.1		Y Y Y G S S Y
BALB/c	Myeloma T15	D	Y Y G S S
C57BL	DFL16 family: Myeloma C3	D	Y Y G S S
C57BL	DFL16 family: Hybridoma 293	D	Y Y G S S
C57BL	DFL16 family: Hybridoma 23169	N	Y Y G S S
C57BL	DFL16 family: Hybridoma 1613	N	Y Y G S *A D Y*
BALB/c	DFL16 family: Myeloma M167	D A	*D* Y G *N* S Y *F G*
BALB/c	DSP2 family: Germline DSP2.5		Y Y G N Y
C57BL	DSP2 family: Hybridoma 2857	D A	Y Y G N Y *G G*
BALB/c	DSP2 family: Germline DSP2.2		Y Y D Y D
BALB/c	DSP2 family: Hybridoma HPCM6	D	Y Y D Y *P*
BALB/c	DSP2 family: Hybridoma HPCM4	D	*F* Y *R* Y D
BALB/c	DSP2 family: Myeloma W3207	N	Y Y *K* Y D
BALB/c	DQ52 family: Germline DQ52		N W D
CBA/J	DQ52 family: Hybridoma 6G6	A	N W D
CBA/J	Hybridoma 6F9	N	*Y Y D G S*
CBA/J	Hybridoma 7C6	D	*Y Y D G S*

* D region germline sequences are translated from nucleotide sequences of Kurosawa and Tonegawa (39). The SP2 family consists of at least eight closely related genes, several of which could be used to generate the observed protein D segment sequences. The germline sequences displayed are those requiring the smallest number of nucleotide substitutions to generate the observed protein sequences. The HPCM4, HPCM6, and W3207 sequences are from Gearhart et al. (6) and the M167 sequence is from Rudikoff and Potter (11). The M167 D_H region has been placed in the DFL16 group based on nucleotide homology as previously proposed (39). Because this D region has an Asn at the fourth position, it could also conceivably be derived from the DSP2 family. Italicized amino acids are not encoded in any known germline D segments.

attributed to the use of alternative recombination sites for the joining of V_H, D_H, and J_H genes. Similar variation in the amino acid sequence and length of this region exists among the PC-binding antibodies of BALB/c mice as well (6, 10, 11). Furthermore, it is interesting that anti-PC antibodies in all three inbred strains studied to date have been found to express Asp or Asn at position 95, the site of V_H-D_H joining.

Several possibilities may explain the asparagine-aspartic acid substitutions at position 95. First, theV_H segments of these proteins may be encoded by two different genes, one that encodes asparagine and one that encodes aspartic acid at position 95. Second, because position 95 is situated at the site of V_H-D_H recombination, one of these amino acids may be encoded by the V_H gene and the other encoded by the D gene. Only a substitution in the first base of the codon for aspartic acid can convert it to one coding for asparagine and vice versa. Therefore, this substitution cannot be a product of a hybrid codon derived from the V_H and D_H gene segments. The T15 V_H gene of BALB/c mice encodes an aspartic acid at position 95 (7, 8). Thus, it is likely that the C3 V_H gene also encodes an aspartic acid at this position, and that the asparagine is derived from the D gene. A third possibility may be that either one of the two amino acids at position 95 is the product of a frequent point mutation occurring within the V_H gene during V_H-D_H rearrangement.

The length of the V_H gene used to encode these H chains may vary as a result of the use of alternate recombination sites for the joining of V_H and D_H genes. If the asparagine at position 95 is encoded by the D gene, then the length of the V_H gene is at least three nucleotides shorter than those coding for H chains with an aspartic acid at this position. An additional alternative recombination site may be used to encode the H chain of 2857 (Fig. 4). This H chain has an alanine immediately following the aspartic acid at position 95. The T15 V_H gene of BALB/c mice, from which most anti-PC V_H regions appear to derive, contains a codon for alanine immediately following the aspartic acid codon, which is not normally expressed. If the C3 V_H gene of C57BL/6N mice also contains an alanine codon at this position, then an alteration in the recombination site of the C3 V_H gene to include at least two base pairs of the alanine codon as part of the V_H gene may have occurred to form the H chain of 2857. This alteration has been proposed to occur in the M167 H chain of BALB/c origin, which also has an alanine at this position, and may also have occurred in the BALB/c PC-binding hybridoma protein HPCG 13 (6). Alternate recombination sites have been identified at the 5′ boundary to J_H genes (7, 35, 36), as well as at the V_K and J_K boundaries (37, 38). However, the extensive differences found within the anti-PC D regions clearly indicate the usage of multiple D genes.

To date, nucleotide sequences have been reported for 10 germline D segments, which fall into three families designated DFL16, DSP2, and DQ52 (39–41). Representatives of these families and the D_H region protein sequences from anti-PC H chains are given in Table II. The same D region sequence is found in proteins 23169, 293, T15, and C3. Although this sequence is not exactly encoded as a germline D segment, it can easily be derived from the DFL16.1 germline sequence and its occurrence in both BALB/c and C57BL points out the conservation of this D family. Similarly, the C57BL 2857 D_H segment can, with the exception of the last two amino acids, be derived from the BALB/c DSP2.5 or DSP2.7 germline genes, which can encode identical D segments, and the CBA/J 6G6 D_H is identical to the BALB/c

DQ52 germline D, which further demonstrates the conservation of D families as was previously observed for both V_K and V_H.

D_H protein segments, such as those found in 1613, 2857, and M167 (Table II), are unusual in that they are longer than those found in most anti-PC H chains and have amino acids at the carboxyl end not encoded in the germline D segments from which they appear to be generated. One possible explanation for these longer D_H segments is that they result from the fusion of two existing germline D genes. Although no germline D genes have been characterized with appropriate spacer configuration on the 5′ and 3′ ends to permit D-D joining, a mechanism has been proposed (39) that would allow the use of D gene coding sequences to substitute for recognition sequences and thus permit D-D fusion. It should be noted, however, that the sequences Ala-Asp-Tyr, Gly-Gly, and Phe-Gly found at the carboxyl end of the D_H segments of proteins 1613, 2857, and M167 are not encoded in any of the D region germline genes in any reading frame. Therefore, it is not readily apparent that any of the PC-D_H segments originate by D-D fusion. A second possible explanation for these sequences is that at the DNA level, nucleotides are inserted randomly during D_H-J_H joining by repair enzymes (i.e., terminal transferase). Even the most diverse of these D_H segments can be theoretically generated by point mutations in existing germline D genes and/or in instances such as M167, the insertion of additional nucleotides at the 3′ end as seen in the following examples (Table II): (*a*) the 1613 segment can be generated from the DFL16.1 gene by the addition of nine nucleotides encoding Ala-Asp-Tyr; (*b*) the 2857 D_H segment can be generated by the addition of four to six nucleotides encoding Gly-Gly (two nucleotides may be contributed by J_H); (*c*) the M167 D_H segment can be generated by two single-base substitutions plus the addition of four to six nucleotides encoding the sequence Phe-Gly; (*d*) the HPCM4 D_H segment can be generated by one single-base and one two-base change in the DSP2.2 gene; (*e*) and the W3207 D_H segment can be generated by a two-base change in the DSP2.2 gene. It is interesting that most amino acids occurring at the carboxyl or 3′ end of D derive from codons high in G-C content. A similar observation has been made for amino acids inserted at the NH_2 or 5′ side of D, the site of V_H-D_H joining (39). It is thus possible that extra nucleotides on either side of D are added by a similar mechanism that possibly involves random addition via repair enzymes. Furthermore, this process appears to occur frequently in H chains, but rarely, if at all, at the site of V_K-J_K joining.

The CBA/J proteins 7C6 and 6F9 have identical D regions and represent the only instance of a D segment other than that of the T15 type appearing in more than one PC-binding antibody. This D segment is probably encoded in the CBA/J germline and may represent the product of a D gene not found in BALB/c or C57BL.

D Region Function. The D region diversity among PC-binding hybridoma and myeloma proteins of BALB/c, C57BL/6N, and CBA/J mice may reflect the relative unimportance of this region to the binding of PC, even though it constitutes a portion of the third hypervariable region. The majority of the amino acids that contact the PC haptenic determinant are in the first and second hypervariable regions of the H chain (42, 43). The only contributions to the binding of PC by the third hypervariable region are relatively weak van der Waal's forces associated with the tryptophan residue at position 100b of the J region. Furthermore, hybridomas with identical V_H segments, such as 2857 and 293, (Fig. 4) but different D regions, have been produced from mice immunized with the same antigen (Table I). Conversely, proteins 23169

and 293 that express the same D region were derived from mice immunized with different antigens. PC-binding hybridoma proteins of BALB/c origin (6) express five different D regions, but were all isolated from mice immunized with the same antigen. The structural diversity observed in these D regions is exemplified by the 6G6 D_H segment (Fig. 4, Table II), which appears to have no relationship to other D segments found in PC-binding antibodies. Thus, the D region does not appear to be involved in either PC binding or PC-carrier specificity.

The D region sequences additionally do not correlate with the L chain present in a given hybridoma protein and therefore must not be required for the binding of PC by given H and L chain combinations or for the association of these particular H and L chains. The C57BL/6N hybridoma proteins 1613 and 23169 (Fig. 4) have H chains with different D regions, but pair with identical L chains (Fig. 1). Proteins 23169 and 293 have H chains with identical D regions that pair with different L chains. The CBA/J hybridoma proteins 6F9 and 7C6 also have H chains with identical D regions that pair with different L chains (Fig. 5). Similar examples exists among the BALB/c PC-binding hybridoma proteins as well (6). Taken together, these observations indicate that D region sequence plays no critical functional role in either antigen binding or V_H-V_L association among anti-PC antibodies. However, it is possible that D region length may be important in the overall structure of the binding site or in effecting V_H-V_L pairing. For example, all antibodies that use the T15 L chain have third hypervariable regions of the same length, although this region in molecules with either M603 or M167 L chains can differ in length by one or two amino acids. It is interesting that, as in the case of the T15 molecules, anti-inulin (44–46), anti-galactan (47), and anti-dextran (48) antibodies have third hypervariable regions (including D) that show no variation in length although considerable primary sequence diversity is found.

J Regions. With the exception of the 6G6 sequence (Fig. 5), all H chains from PC-binding antibodies have been found to use the J_H1 segment. The reason for the exclusive use of the J_H1 gene is unclear. The only J_H contribution to the binding of the PC haptenic determinant are van der Waal's forces associated with the tryptophan at position 100b. Both the J_H1 and J_H3 genes of BALB/c mice encode a tryptophan at this position. Furthermore, the recombination of V_H, D_H, and J_H has deleted this tryptophan from MOPC-167 of BALB/c origin and 2857 of C57BL/6N origin, which implies that this residue is not essential for PC binding. It also seems unlikely that the J_H1 region is required for interactions with the carrier portion of the antigenic molecule because, regardless of the immunogen used, the hybridoma proteins produced express the J_H1 region, which argues against its involvement in PC-carrier specificity. Alternatively, a mechanism may exist for joining specific V_H, D_H, and J_H genes such that the T15 and C3 V_H genes preferentially recombine with the J_H1 gene. The 6G6 H chain, which uses a different V_H gene, expresses the J_H3 region. Interestingly, MOPC-47, the only BALB/c protein homologous to 6G6 for which a complete H chain variable region amino acid sequence is available, also expresses the J_H3 region (49).

The J_H1 sequence used in the C57BL H chains differs from that of the BALB/c J_H1 by the occurrence of Thr in place of Gly at position 105 in all C57BL J_H1 segments (Fig. 4). Because this substitution is present in all four H chains, this J_H region must be germline encoded and probably represents an allelic form (J_H1^b) of the J_H1 gene

of BALB/c mice (J_H1^a). The phenotypes of anti-PC antibodies expressed in CBA/J are unusual in that this strain uses a C57BL-like PC-V_H gene and a BALB/c J_H1 segment. The origin of the CBA/J haplotype and its relation to BALB/c and C57BL is unclear as documentation concerning the derivation of these strains is at best ambiguous (50). The observation that BALB/c and CBA/J mice share many IgC_H allotypic determinants, but that CBA/J mice express the C57BL PC-V_H gene, suggests that at least two haplotypes, as represented by the BALB/c IgC_H and C57BL IgV_H loci, have been involved in the derivation of CBA/J. The CBA/J immunoglobulin phenotype indicates a possible crossover between the BALB/c IgC_H locus and the C57BL V_H locus that initially results in the phenotype V_H^b-J_H1^a-C_H^a. After this event, additional changes may have occurred in the C_H complex that result in the present phenotype V_H^b-J_H1^a-C_H^j (15).

Additional insights into the events and mutational processes involved in the generation of antibody diversity will probably be obtained by similar comparative studies involving other inbred strains of mice. Furthermore, the extension of these investigations to wild mice species at both the protein and DNA levels will provide an opportunity to assess the evolution of these particular gene families through the genus *Mus*.

Summary

Complete variable (V) region amino acid sequences were determined for four heavy (H) and one light (L) chain from C57BL phosphocholine (PC)-binding monoclonal antibodies. Additional NH_2-terminal sequences were obtained from H and L chains of C57BL and CBA/J origin. When these V regions were compared with previously reported anti-PC sequences, a number of observations could be made regarding the function and evolution of L and H chain segments used in these antibodies. (*a*) L and H chain V segments are remarkably conserved in these inbred strains, although there has been an accumulation of point mutations identifying apparently allelic forms of V_K and V_H. (*b*) Mice of each genotype use the same three V_K segments in combination with a single V_H segment to produce most anti-PC antibodies. An exception has been noted that indicates the occasional use of a second V_H gene segment. (*c*) Multiple, different D_H regions are used by mice of each strain, which suggests that the D_H segment sequence plays no critical role in either antigen binding or V_H-V_L pairing. Furthermore, the D_H segments and their corresponding gene families appear to be highly conserved in the inbred strains studied. (*d*) Most PC-binding antibodies use the J_H1 joining segment. All J_H1 sequences from C57BL mice differ from the BALB/c J_H1 at position 105, which identifies allelic forms of the J_H1 region. These studies are a first assessment of the nature of mutational events associated with the evolution of specific multigene immunoglobulin families and indicate that homologous V_H, D_H, J_H, V_K, and J_K genes are similarly assembled and expressed in PC antibodies from three diverse genotypes.

Note added in proof: Alt and Baltimore (*Proc. Natl. Acad. Sci. U. S. A.*, **79**:4118) have similarly suggested that additional nucleotides found at the borders of D are inserted by terminal transferase.

Received for publication 26 July 1982.

References

1. Barstad, P., S. Rudikoff, M. Potter, M. Cohn, W. Koningsberg, and L. Hood. 1974. Immunoglobulin structure: amino terminal sequences of mouse myeloma proteins that bind phosphorylcholine. *Science (Wash. D. C.).* **183:**962.
2. Hood, L. E., D. McKean, V. Fansworth, and M. Potter. 1973. Mouse immunoglobulin chains. A survey of the amino terminal sequences of K chains. *Biochemistry.* **12:**741.
3. Potter, M. 1977. Antigen-binding myeloma proteins of mice. *Adv. Immunol.* **25:**14.
4. Kocher, H. P., C. Berek, M. H. Schrier, H. Cosenza, and J. C. Jaton. 1980 Phosphorylcholine-binding hybridoma proteins of normal and idiotypically suppressed BALB/c mice. II. Variable region N-terminal amino acid sequences. *Eur. J. Immunol.* **10:**264.
5. Claflin, J. L., S. Hudak, and A. Maddalena. 1981. Anti-phosphorylcholine hybridoma antibodies. I. Direct evidence for three distinct families of antibodies in the murine response. *J. Exp. Med.* **153:**352.
6. Gearhart, P., N. D. Johnson, R. Douglas, and L. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.).* **291:**29.
7. Early, P. W., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. *Cell.***19:**981.
8. Crews, S., J. Griffin, H. Huang, K. Calame, and L. Hood. 1981. A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell.* **25:**59.
9. Clarke, S. H., J. L. Claflin, and S. Rudikoff. 1982. Polymorphisms in immunoglobulin heavy chains suggesting gene conversion. *Proc. Natl. Acad. Sci. U. S. A.* **79:**3280.
10. Rudikoff, S., and M. Potter. 1974. Variable region sequence of the heavy chain from a phosphorylcholine binding myeloma protein. *Biochemistry.* **13:**4033.
11. Rudikoff, S., and M. Potter. 1976. Size differences among immunoglobulin heavy chains from phosphorylcholine-binding proteins. *Proc. Natl. Acad. Sci. U. S. A* **73:**2109.
12. Rudikoff, S., and M. Potter. 1980. Allelic forms of the immunoglobulin heavy chain variable region. *J. Immunol.* **124:**2089.
13. Claflin, J. L., and S. Rudikoff. 1979. Structural evidence for a polymorphic or allelic form of the heavy chain variable region. *J. Immunol.* **122:**1402.
14. Lieberman, R., S. Rudikoff, W. Humphrey, Jr., and M. Potter. 1981. Allelic forms of anti-phosphorylcholine antibodies. *J. Immunol.* **126:**172.
15. Lieberman, R. 1978. Genetics of I_GC_H (allotype) locus in the mouse. *In* Springer Seminar in Immunopathology. P. A. Miesher and H. J. Mueller-Eberhard, editors. Springer-Verlag, New York. **1:**7–30.
16. Shulman, M., C. D. Wilde, and G. Kohler. 1978. A better cell line for making hybridomas secreting specific antibodies. *Nature (Lond.).* **276:**269.
17. Pawlita, M., E. Mushinski, R. J. Feldman, and M. Potter. 1981. A monoclonal antibody that defines an idiotype with two subsites in galactan-binding myeloma proteins. *J. Exp. Med.* **154:**1946.
18. Chesbro, B., and H. Metzger. 1972. Affinity labelling of a phosphorylcholine binding mouse myeloma protein. *Biochemistry.* **11:**766.
19. Bridges, S. H., and J. R. Little. 1971. Recovery of binding activity in reconstituted mouse myeloma proteins. *Biochemistry.* **10:**2525.
20. Wittman-Liebold, B., H. Graffunder, and H. Kohls. 1976. A device coupled to a modified sequenator for the automated conversion of anlinothiazolines into PTH derivatives. *Ann. Biochem.* **75:**621.
21. Hunkapillar, M. W., and L. Hood. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (Polybrene) and high pressure liquid chromatography. *Biochemistry* **17:**2124.

22. Zimmerman, C. L., E. Appella, and J. J. Pisano. 1977. Rapid analysis of amino acid phenylthiohydantoins by high performance liquid chromatography. *Anal. Biochem.* **77:**569.
23. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T 4. *Nature (Lond.).* **227:**680.
24. Jolley, M. E., and C. P. J. Glaudemans. 1974. The determination of binding constants for binding between carbohydrate ligands and certain proteins. *Carbohydr. Res.* **33:**377.
25. Kohler, G., and C. Milstein. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature (Lond.).* **256:**495.
26. Andres, C. M., A. Maddalena, S. Hudak, N. M. Young, and J. L. Claflin. 1981. Anti-phosphorylcholine hybridoma antibodies. II. Functional analysis of binding sites within three antibody families. *J. Exp. Med.* **154:**1584.
27. Metzger, H., B. Chesbro, N. M. Hadler, J. Lee, and N. Otchin. 1971. Modification of immunoglobulin combining sites. *Prog. Immunol. 1st Congr. Immunol.* 254–266.
28. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1979. Sequence of Immunoglobulin Chains. National Institutes of Health Publication 80-2008. 1–183.
29. Selsing, E., and U. Storb. 1981. Somatic mutation of immunoglobulin light chain variable region genes. *Cell.* **25:**47.
30. Max, E. E., J. G. Siedman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin K constant region gene. *Proc. Natl. Acad. Sci. U. S. A.* **76:**3450.
31. Claflin, J. L. 1976. Uniformity in the clonal repertoire for the immune response to phosphorylcholine in mice. *Eur. J. Immunol.* **6:**669.
32. Claflin, J. L., and M. Cubberly. 1980. Clonal nature of the immune response to phosphorylcholine. VII. Evidence throughout inbred mice for molecular similarities among antibodies bearing the T15 idiotype. *J. Immunol.* **125:**551.
33. Williams, K. R., and J. L. Claflin. 1980. Clonotypes of anti-phosphorylcholine antibodies induced with *Proteus morganii* (Potter). I. Structural and idiotypic similarities in a diverse repertoire. *J. Immunol.* **125:**2429.
34. Wolfe, J., and J. L. Claflin. 1980. Clonal nature of the immune response to phosphorylcholine. IX. Heterogeneity among antibodies bearing M511 idiotypic determinants. *J. Immunol.* **125:**2397.
35. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy chain genes. *Nature (Lond.).* **286:**676.
36. Gough, N. M., and O. Bernard. 1981. Sequences of the joining region genes for immunoglobulin heavy chains and their role in generation of antibody diversity. *Proc. Natl. Acad. Sci. U. S. A.* **78:**509.
37. Weigert, M., R. Perry, D. Kelley, T. Hunkapiller, J. Schilling, and L. Hood. 1980. The joining of V and J gene segments creates antibody diversity. *Nature (Lond.).* **283:**497.
38. Rudikoff, S., D. N. Rao, C. P. J. Glaudemans, and M. Potter. 1980. K chain joining segments and structural diversity of antibody combining sites. *Proc. Natl. Acad. Sci. U. S. A.* **77:**4270.
39. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* **155:**201.
40. Sakano, H., Y. Kurosawa, M. Weigert, and S. Tonegawa. 1981. Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy chain genes. *Nature (Lond.).* **290:**562.
41. Kurosawa, Y., H. von Boehmer, W. Haas, H. Sakano, A. Trauneker, and S. Tonegawa. 1981. Identification of D segments of immunoglobulin heavy chain genes and their rearrangement in T lymphocytes. *Nature (Lond.).* **290:**565.
42. Segal, D. M., E. A. Padlan, G. C. Cohen, S. Rudikoff, M. Potter, and D. R. Davies. 1974.

The three dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. U. S. A.* **71:**4298.

43. Padlan, E. A., D. R. Davies, S. Rudikoff, and M. Potter. 1976. Structural basis for the specificity of phosphorylcholine-binding immunoglobulins. *Immunochemistry.* **13:**945.
44. Vrana, M., S. Rudikoff, and M. Potter. 1978. Sequence variation among heavy chains from inulin binding myeloma proteins. *Proc. Natl. Acad. Sci. U. S. A.* **75:**1957.
45. Rudikoff, S., and M. Potter. 1981. Immunoglobulin heavy chains from anti-inulin myeloma proteins: evidence for a new heavy chain joining segment. *J. Immunol.* **127:**191.
46. Johnson, N., J. Slankard, L. Paul, and L. Hood. 1982. The complete V domain amino acid sequences of two myeloma inulin-binding proteins. *J. Immunol.* **128:**302.
47. Rao, D. N., S. Rudikoff, H. Krutzsch, and M. Potter. 1979. Structural evidence for independent joining region gene in immunoglobulin heavy chains from anti-galactan myeloma proteins and its potential role in generating diversity in complimentarity determining regions. *Proc. Natl. Acad. Sci. U. S. A.* **76:**2890.
48. Schilling, J., B. Clevinger, J. M. Davie, and L. Hood. 1980. Amino-acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V region gene segments. *Nature (Lond.).* **283:**35.
49. Robinson, E. A., and E. Appella. 1977. Amino acid sequence of the first 217 residues of a mouse heavy chain (MOPC 47A) with a domain deletion. *Proc. Natl. Acad. Sci. U. S. A.* **74:**2465.
50. Altman, P. L., and D. D. Katz, editors. 1979. Inbred and genetically defined strains of laboratory animals. Part I. Mouse and rat. *Biol. Handb.* **III:**1.