DIVERSITY IN THE GERMLINE ANTIBODY REPERTOIRE

Molecular Evolution of the T15 V_H Gene Family

By ROGER M. PERLMUTTER,* BENNETT BERSON, JOHANNA A. GRIFFIN, AND LEROY HOOD

From the *Departments of Medicine and Biochemistry, and the Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195; and the Division of Biology, California Institute of Technology, Pasadena, California 91125

In vertebrates, active immunity depends in part upon the ability to generate a heterogeneous repertoire of antigen-binding structures encoded by elements of a superfamily of antibody and T cell receptor genes. The antibody gene superfamily includes at least six sets of elements that undergo rearrangement in lymphoid cells: those encoding antibody heavy chains, the κ and λ antibody light chains, and the α, β, and γ gene families of the T cell receptor (1–6). In each case, with the possible exception of the γ genes, a limited repertoire of germline gene segments can be used to generate a much larger repertoire of recognition structures through combinatorial joining of these segments before transcription and combinatorial association of the mature polypeptide chains of the heterodimeric receptor after translation.

Considerable progress has been made in cataloging the germline elements of these families and in defining the involvement of these elements in specific immune responses. This is especially true of antibody heavy chain variable region $(V_H)^1$ gene segments, which in the mouse have been assigned to eight distinct, noninterspersed V_H gene families containing at least two (V_HX24) and in some cases more than 50 (V_HJ558) elements, based upon sequence homology (7). Combinatorial mechanisms notwithstanding, antibody heterogeneity frequently results from somatic mutational processes acting upon one or at most a few germline V_H gene segments. Thus, antibodies binding glutamic acid-alanine-tyrosine (8), nitrophenacetyl haptens (9), group A streptococcal carbohydrate (10), and, especially, phosphorylcholine (PC) (11) all adhere to this formula. Since the preimmune B cell repertoire appears to include a random representation of V_H, diversity (D), and heavy chaining joining region (J_H) gene segments, a stochastic model for generation of the antibody repertoire has been proposed (12) that posits selection of appropriate combining sites by antigen-induced proliferation. Somatic mutation would then operate to permit further selection of variant antibodies on the basis of increased affinity (12, 13).

Under such a scheme it might well be imagined that frequently encountered environmental antigens would impose significant selection pressure on the evo-

^{1} Abbreviations used in this paper: C region, constant region; D region, diversity region; J region, joining region; PC, phosphorylcholine; V_H region, heavy chain variable region.

lution of the germline variable region repertoire. Somatic mutation and subsequent selection would function best when operating on a substrate that already has high affinity for cognate antigen. In this way the entire germline repertoire of V_H gene segments has presumably been fixed in response to the universe of significant pathogens.

At the same time, three observations support the proposition that antibody V gene segments may not be subject to strong evolutionary pressure. First, mice suppressed for κ light chain production and hence able to use only two V_λ gene segments nevertheless respond satisfactorily to antigenic challenge (14). Secondly, the murine V_H families contain large numbers of pseudogenes, typically 25–30% (1, 7), indicating that even deleterious mutations can be fixed at a high rate in these genes. Finally, and more persuasively, the rate of nucleotide substitution within germline V_H gene segments appears to be very high. For example, Loh et al. (15) compared the sequences of five NP^b-related V_H gene segments from BALB/c DNA with their C57BL10 counterparts and were unable to define allelic relationships among these elements, presumably because point mutations and recombination or conversion events had obscured the origin of each individual gene segment such that these sequences could not be separated into pairs clearly sharing a common precursor.

With these thoughts in mind, we sought to examine the forces that modify the germline antibody gene repertoire. The T15 V_H gene family provides an ideal model for this analysis since, in the BALB/c mouse, it includes only four elements, all of which are >88% homologous (16). One of these elements (V1) encodes virtually all PC-binding heavy chains in BALB/c mice (11). Because PC is a determinant on a large number of bacterial pathogens (17), and since antibodies to PC can be shown to protect against in vivo challenge with viable *Pneumococcal* organisms in mice (18), there is reason to believe that the V1 gene segment is important in generating an effective immune defense and should therefore be subject to relatively strong environmental selection pressure. A second T15 family member, V3, is a pseudogene (19) and thus, though 89% homologous with V1, should drift in the absence of significant selection.

To test these conjectures, we have examined the T15 V_H gene family in a number of inbred mouse strains and have defined for the first time V_H alleles by DNA sequence analysis. That is, we have been able to identify pairs of V_H sequences that must have shared a common precursor before the divergence of BALB/c and C57BL/10 mouse strains. Our data reveal that individual V_H gene segments within the same V_H gene family evolve rapidly and at different rates, and support a model for the generation of the germline antibody repertoire through frequent gene duplication and conversion events.

Materials and Methods

Genomic Blotting. High molecular weight genomic DNA was isolated from liver or sperm of individual mice (The Jackson Laboratory, Bar Harbor, ME) or Lewis rats (maintained at California Institute of Technology) according to the method of Blin and Stafford (20). Sperm from B10.P mice (C57BL/10 animals congenic for H-2^p) were obtained from Dr. Jeffrey Frelinger (University of North Carolina). 10 μg of each genomic DNA was completely digested with restriction enzyme (Bethesda Research Laboratories, Gaithersburg, MD) before electrophoresis in 0.7% agarose in TAE buffer (0.04 M Tris

acetate, pH 7.5, 0.002 M EDTA) for 14 h at 40 V. Blotting of DNA fragments and hybridization were performed as previously described (10). The probe used was pS107V1, a pBR322 cDNA clone containing the leader and V$_H$ coding region from the BALB/c myeloma S107 which uses a rearranged V1 gene identical to the germline gene segment throughout (16).

Library Construction. B10.P sperm DNA was partially digested with *Mbo* I (New England Biolabs, Beverly, MA) (0.1 U/µg for 15 min) and fragments with an average size of 20 kb were purified by sucrose gradient centrifugation (21). These inserts were ligated into the *Bam* HI site of the L47.1 phage vector (22), packaged in vitro, and plated on KH802 host cells as previously described (10). Additional libraries were constructed using DNA fragments eluted from low melting point agarose gels (Bethesda Research Laboratories) cloned into L47.1 phage arms.

Library Screening. Libraries containing 10^6 phage were screened using the pS107V1 plasmid labeled to a specific activity of 10^8 cpm/µg by nick translation (23) using ^{32}P-nucleotide triphosphates (New England Nuclear, Boston, MA). Positive colonies were identified on duplicate filters, picked, rescreened, and grown in liquid culture for later analysis.

DNA Sequence Analysis. Appropriate restriction fragments were ligated into the M13mp8 bacteriophage vector (Collaborative Research, Inc., Lexington, MA) for sequencing using the dideoxy chain termination method (24) with a 12 basepair (bp) primer (Pharmacia-PL Inc., NJ). Clones were obtained using *Sau* 3a, *Alu* I, or *Rsa* I digests and were selected by hybridization to the ^{32}P-labeled S107 probe. Sequencing was performed using either ^{32}P- or ^{35}S-labeled nucleotides (New England Nuclear). Additional 3′ flanking region sequencing was performed using a specific oligonucleotide (5′-GGACAGTGCC-ACTTA-3′) as primer.

Results

T15 V$_H$ Family Has Undergone Recent Deletion and Expansion Events. Considerable evidence suggests that all mice contain V$_H$ sequences homologous to the BALB/c V1 sequence. In particular, a binding site–related idiotype is shared by most murine strains and species (25), all of which respond to PC, and amino acid sequence data confirms a close relationship for PC-binding heavy chains from BALB/c, C57BL/6, and A/J mouse strains (26) and even for rat PC-binding heavy chains (11). This relationship can be visualized in Fig. 1, which shows the T15 gene family in eight mouse strains representing the major IgCH allotype groups (27) and in Lewis rats as visualized by genomic blotting using a T15 V$_H$ probe. The number of elements identified by this probe varies considerably in the different DNA samples. For example, the BALB/c DNA contains four prominently hybridizing *Eco* RI restriction fragments, representing the four members of the T15 gene family as previously reported (16), while A/J DNA contains only two of these fragments (corresponding in size to the V1 and V13 bands of BALB/c on an *Eco* RI digest) and the wild mouse DNA appears to contain six strongly hybridizing species. More dramatically, rat DNA contains at least 14 strongly hybridizing bands, representing a series of gene duplications (or deletions in the mouse) that must have occurred during the relatively short evolutionary interval (<10 million years [28]) separating these two rodent species. Although assignment of allele designations to particular bands on a genomic blot is dangerous (see below), in general the V1 and V13 bands appear most closely conserved on these blots.

Cloning of the B10.P T15 V$_H$ Gene Family. Initial screening of a B10.P sperm DNA library (which included 10^6 primary recombinants and had been amplified
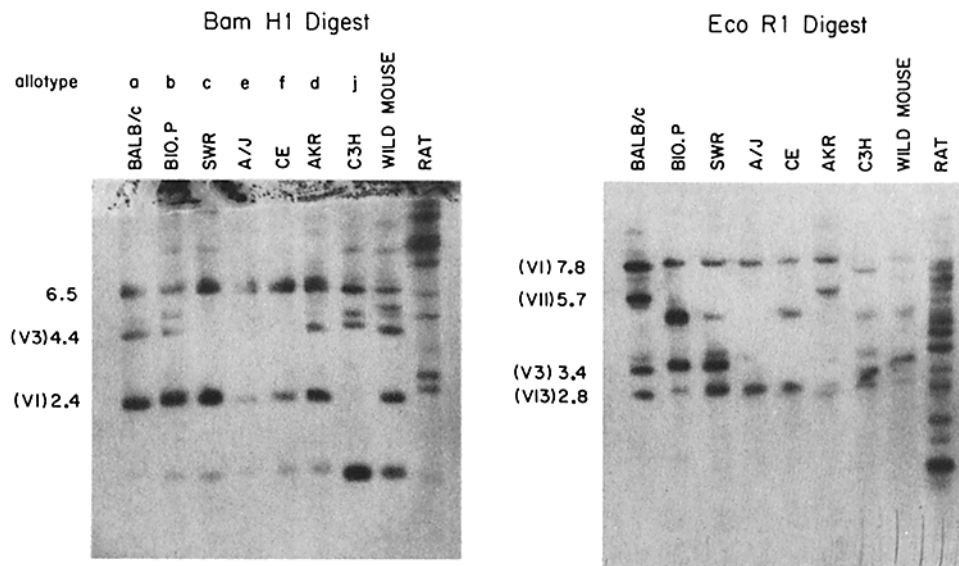
FIGURE 1. A Southern blot analysis of the T15 V_H gene family in rodents. Shown are the patterns obtained when *Eco* RI– or *Bam* HI–digested DNA from the indicated strains is examined by genomic blotting using the pS107V1 probe (16). The heavy chain allotype group (27) is also listed for each mouse strain, as are the sizes (kb) and identities of the major hybridizing restriction fragments in BALB/c DNA.

once) with the pS107V1 probe yielded seven examples of phage λVB2.4 and 14 examples of phage λVB4.4 that contained inserts corresponding, respectively, to the 2.4 and 4.4 kb *Bam* HI fragments seen in Fig. 1. Comparison of *Sau* 3a digests indicated that multiple copies of only two genomic clones had been obtained, probably reflecting selection of these particular phages during amplification of the library. To obtain a more complete representation of the T15 V_H family in B10.P mice, B10.P sperm DNA was completely digested with *Bam* HI and fragments of average size 6.5 kb were isolated by sucrose gradient centrifugation (21). A library was then constructed using these fragments and L47.1 phage arms, and 10^5 phage were screened using the pS107V1 probe. Nine positive clones were identified including seven with a 6.5 kb *Bam* HI insert (λVB6.5), one with a 4.8 kb *Bam* HI insert (λVB4.8), and another example of λVB4.4 that had been obtained previously. Thus, clones corresponding to all four strongly hybridizing *Bam* HI genomic fragments from B10.P DNA were obtained. Partial restriction maps and sequencing strategies for these four elements are shown in Fig. 2.

Sequence Analysis of the B10.P T15 V_H Gene Family. The complete V_H sequences for the four members of the T15 V_H family in B10.P mice are shown in Fig. 3 beginning 35 bp 5′ to the translation start codon, extending through the leader, intervening sequence, V_H-coding region, and the 3′ canonical rearrangement signals (underlined), and including 51 bp of the 3′ flanking regions. Alignment of these sequences was achieved by visual comparison and, although significant gaps must be inserted within the intron and flanking region segments, the V_H-coding regions differ by at most 41 of 312 (13%) substitutions for λVB2.4
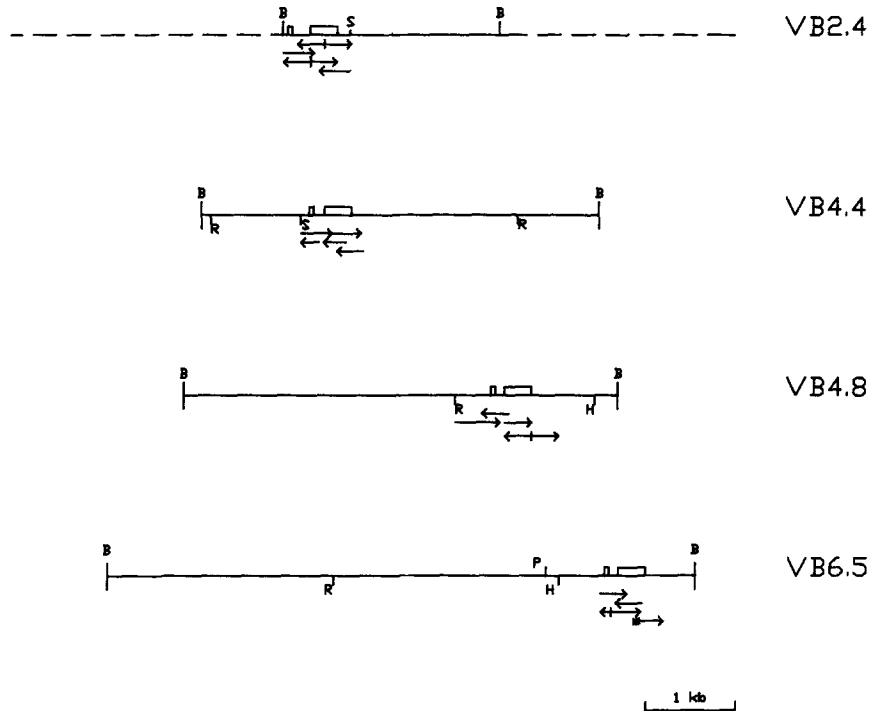
VB2.4

VB4.4

VB4.8

VB6.5

1 kb

FIGURE 2. Sequencing strategies used to examine λ phage clones containing the four elements of the B10.P T15 V$_H$ gene family. The identity of the phage clone is shown at right. Dashed lines indicate additional unmapped genomic DNA. The restriction sites shown are: B, *Bam* HI; H, *Hind* III; P, *Pst* I; R, *Eco* RI; S, *Sau* 3a. An asterisk marks a sequence primed with a specific oligonucleotide (see text for description). The scale is shown at bottom right.

compared with λVB4.8. Thus, the B10.P T15 V$_H$ gene family, like its BALB/c counterpart, consists of four closely homologous elements.

Defining Alleles within the T15 V$_H$ Gene Family. To assign the proper BALB/c allele designations for members of the B10.P T15 V$_H$ family, we first compared the nucleotide sequences of each member of the BALB/c family with each member of the B10.P family. The results of these comparisons are presented in a percent homology matrix in Table I. By this method, only λVB2.4 can be identified in that it is clearly more homologous to BALB/c V1 than to any other BALB/c sequence. Both λVB4.4 and λVB6.5 were 97% homologous to BALB/c V11 while λVB4.8 was 95% homologous to both V11 and V13 of BALB/c. The levels of homology were thus so similar between these sequences that reliable allele designations cannot be proposed. With the exception of V1, which can be defined by sequence homology, additional criteria are required to unambiguously identify alleles within the T15 V$_H$ gene family. The definition of each allelic pair is outlined below.

V1B10 Gene Segment Encodes PC-binding Heavy Chains from Ig-1^b Allotype Mice. Defined by nucleotide sequence homology (Table I), four replacement substitutions and one silent substitution separate the λVB2.4 gene segment of B10.P (Ig-1^b) mice and the V1 gene segment of BALB/c (Ig-1^a) mice: three transitions and two transversions. The translated amino acid sequence of the

λVB2.4 gene segment is identical with that determined by Clarke et al. (26) for four PC-binding heavy chains of C57BL/10 origin (Fig. 4). Thus, just as the V1 gene segment encodes most PC-binding heavy chains in BALB/c mice, the λVB2.4 gene segment encodes PC-binding heavy chains in C57BL/10 mice. In addition, these two gene segments are homologous throughout the leader, intervening sequence, and 5′ flanking regions, and different in these regions from all other members of the T15 V_H gene families of either mouse strain (Fig. 3, and G. Siu and R. M. Perlmutter, unpublished results). Clearly, V1BALB and λVB2.4 are structurally and functionally allelic.

λVB4.4 Insert Is the Allele of BALB/c V3. In BALB/c DNA the V3 gene segment has been linked 16 kb upstream of the V1 gene segment (11). Two members of the B10.P T15 V_H gene family are also linked, the λVB2.4 segment, allelic to V1BALB, and the λVB4.4 gene segment. Fig. 5 demonstrates that these two elements were isolated on a single λL47.1 phage clone (λ40.2) from the initial B10.P genomic library screen. Thus, λVB4.4 is by position at least the allele of the BALB/c V3 gene segment.

Four aberrancies render V3 a pseudogene in BALB/c mice: (*a*) a 7 bp deletion in the spacer between recognition signals probably destroys the rearrangement capability of this gene segment since considerable evidence supports the essential role of a fixed 12:23 bp spacer ratio in antibody gene superfamily rearrangements (1–6); (*b*) a 4 bp insertion at codon 90 destroys the correct reading frame, resulting in a nonsense mutation; (*c*) an amber codon halts translation at position 47; and (*d*) the absence of satisfactory leader and promoter sequences makes transcription improbable (G. Siu, H. V. Huang, S. T. Crews, and L. Hood, unpublished results). As shown in Fig. 3, λVB4.4 also has a 7 bp deletion within the rearrangement spacer. In this sequence, however, the insertion at codon 90 is not present and codon 47 encodes tryptophan as in other functional family members. Thus, in most respects, λVB4.4 encodes a functional V_H domain, although the 3′ spacer deletion likely precludes DNA rearrangement.

An additional impediment to expression of the λVB4.4 V_H gene segment results from truncation of the normal leader sequence. This is illustrated by the sharp break in homology between λVB4.4 and λVB4.8 beginning at nucleotide position 66 of Fig. 3. Although a methionine codon is positioned in the appropriate reading frame at positions 42–44, examination of the 5′ flanking region sequence reveals no satisfactory promoter elements (CAAT, TATA, or consensus octamer homologies [29, 30]) typical of V_H gene segments (R. M. Perlmutter, unpublished data). An identical defect is also seen in the BALB/c V3 gene segment (G. Siu, H. Huang, S. T. Crews, and L. Hood, unpublished data). Thus, two features of the λVB4.4 gene segment shared with V3BALB render this element a pseudogene in B10.P mice: the absence of satisfactory promoter and leader sequences and inappropriately spaced 3′ rearrangement signals. No substitutions in this gene segment could therefore have been exposed to selective pressure during the period when BALB/c and C57BL/10 chromosomes 12 diverged.

λVB4.4 differs from V3BALB by 15 substitutions (counting the 4 bp insertion in the BALB/c gene at codon 90 as a single event). Perhaps surprisingly, all of

```
                                                                                                      94
λVB2.4  TTCACT-ATCAGAAACAGCACTCAGTCCTGTCACT ATG AAG TTG TGG TTA AAC TGG GTT TTT CTT TTA ACA CTT TTA CAT GCTAATTTATGGGC
λVB4.4  CGATG.GCAT..GG...GT..A..AG..........  GAC TTC A.. TT AAC .T. ATT A.C ..C ACA G.. .C. ... A.. ........G...TG
λVB4.8  ....G.G.....CCTTGAACACAGA-.........C  ... ... G.G ... ... A.. ... G.. G.. C.A ... A... .........G
λVB6.5  G.....ACCTTGAACACAGA-..........C      ... ... C.G ... ... A.. ... G.. G.. ..A ... A... ......T.
```

```
                                                                                                      204
λVB2.4  AAAAACAGTTCCTGTGTGTGTGTGTGTAT------------------------GTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTG
λVB4.4  ...T.C.A.G.AGA..A....A.....--CAAGTCAGAGGACAGTAGATCTGTGACATTTCC--------------------
λVB4.8  ..C...G.A.G..GAA.AC...A....C..G.AAGTAAGAGGACAGTGGATCTGTGACATTTCC----------------
λVB6.5  ..C...G.A.G..GA..A....A.......G.AAGTAAGAGGACAGTGGATCTGTGACATTTCC------------
```

```
                                                                                                      306
λVB2.4  TGTGTGTGTGTGTGTTGTGTTGTGCCATATGTGTAACAGGGATATAGTGCTGTATGACGATGTGTATGACCAGCCTTTCTCTACTATTGCAGGT ATC CAG TCT GAG GTG AAG
λVB4.4  --------------------------------------------------.C...G--.....A.........
λVB4.8  --------------------------------------------------....AA-...T...........
λVB6.5  --------------------------------------------------...T..AA-....T..........  -3
```

```
                                                                                                      387
λVB2.4  CTG GTG GAA TCT GGA GGA GGC TTG GTA CAG TCT GGG CGT TCT CTG AGA CTC TCC TGT GCA ACT TCT GGG TTC ACC TTC AGT
λVB4.4  ... ... ..G ... .A. ... ... ... ... C.. ... G.. ... ... ... ... ... G.. ... ... ..A ... ... ..T .C.
λVB4.8  ... ... ..G ... ... ... ... ... ... C.. ... G.. ... ... ..T ... ... G.. ... ... ..A ... ... ... .C.
λVB6.5  ... ... ... ... ... ... ...
                    10                      20                              30
```

FIGURE 3. Sequence analysis of the B10.P T15 V$_H$ gene family. Shown are the sequences of all four family members beginning 35 bp 5' to the initiation codon. The complete sequence of λVB2.4 is indicated; only sequence differences are displayed for the other three gene segments. Hyphens were inserted to maximize alignment. The nucleotide positions are numbered to the right of each set, and the codon positions are numbered sequentially below the sequence, with codon 1 corresponding to the first amino acid of the mature secreted heavy chain. The heptamer and nanomer recognition sequences are underlined.

```
                                                                                                    40                      50                              478
λVB2.4  GAT TTC TAC ATG GAG TGG GTC CGC CAA GCT CCA GGG AAG GGA CTG GAG TGG ATT GCT GCA AGT AGA AAC AAA GCT AAT GAT
λVB4.4  ... ... .A. ... A.C ... ... T.. ... ..G C.. ... A.. ... .C. ... ... T.G TTT .T. ... ... ... ... ... .G.
λVB4.8  ... ... .A. ... AGC ... ... ... ... ..G C.. ... ... ... .C. ... ... T.G TTG .T. ... ... ... ... ... .G.
λVB6.5  ... ... .A. ... AGC ... ... ... ..G CT. ..A .C. ... ..T T.G .G. TTT .T. ... ... ... ... ... .G.

                                          60                      70                      80                              559
λVB2.4  TAT ACA ACA GAG TAC AGT GCA TCT GTG AAG GGT CGG TTC ATC GTC TCC AGA GAC ACT TCC CAA AGC ATC CTC TAC CTT CAG
λVB4.4  ..C ... ... ..C ... ..T .C. ... A.. ..I ... ... ... .C. A.. ... ... ..T .A. ... ... ... ... ... ..T ..A
λVB4.8  ..C ... ... ... ..T .C. ... ..T ... ... ... .C. A.. ... ... ..T .A. ... ... .A. ... ... ..T ..A
λVB6.5  ..C ... ... ... ... ... ... ... ... ... ... .C. A.. ... ... ..T .A. ... ... ... ... ... ..T ..A

                                          90                      100                                                     650
λVB2.4  ATG AAT CCC CTG AGA GCT GAG GAC ACT GCC ATT TAT TAC TGT GCA AGA GAT GCACACAGTGAGAGGAGGACGTCATTGTGAGCCCAGACACAA
λVB4.4  ... ..C A.A ... ... A.. ... .G. ... ... .C. ... ... ... ... ..A ... A.. A.......... G.------ ... TT.A...
λVB4.8  ... ..C A.. ... ... .G. ... .C. ... .A. ... ... .T. :A. :C. ... ... .T. .T........ G.T.T...G ... T.A.A...
λVB6.5  ... ... ... ... ..G ... .G. ... .C. ... ... .A. ... ...A. -......... G.T.T...G ... ..T...

                                                                                            706
λVB2.4  ACCTCCATTGCCATTGCCAGGGGTGTTC-TGGACCAACAGAGGCGC-ACATGACTATGGAAAA
λVB4.4  ........-.........C..C..AA......CAG.GC.AATGACCACAG.AA..GTGC
λVB4.8  ..........C...-A.......G..GCGC......C.GA.....
λVB6.5  ........-......AA...C...-A...........G..GTGC.......C.GA.....
```

TABLE I

Percentage Homologies Among T15 V$_H$ Family Alleles

		B10.P		
	λVB2.4	λVB4.4	λVB4.8	λVB6.5
V1	98	89	88	88
V3	88	95	93	92
BALB/c				
V11	90	97	95	97
V13	86	92	95	95

these substitutions would lead to amino acid replacements if this gene segment were expressed.

Defining V11 and V13 Alleles in the T15 V$_H$ Gene Family. The V1 and V3 alleles of B10.P mice were easily identified by close homology with their BALB/c alleles and by the unique structural features of the V3 pseudogene in both mouse strains (see above). The remaining two B10.P coding region sequences were not easily classified as alleles of V11 or V13. Table I shows that λVB4.8 is 95% homologous to both V11 and V13 of BALB/c, while λVB6.5 is 97% homologous to BALB/c V11 and 95% homolgous to BALB/c V13. Both of these B10.P sequences appear to represent functional V$_H$-coding regions. Compared with other members of the T15 V$_H$ gene family, λVB6.5 contains a 2 bp deletion adjoining the heptamer recognition sequence (Fig. 3), a feature that is also present in BALB/c V11. Comparison of flanking region and intron sequences indicates that λVB4.8 is most like BALB/c V13 (R. M. Perlmutter, G. Siu, and L. Hood, unpublished data). Based on a unique 2 bp deletion and high overall homology, λVB6.5 is the B10.P allele of BALB/c V11, while λVB4.8 can be classified as the V13 allele because of extensive flanking region homology between these gene segments.

Although the B10.P alleles of all four BALB/c T15 V$_H$ family gene segments have thus been isolated, another homologous band at 2.8 kb is seen on *Eco* RI genomic blots of B10.P DNA (Fig. 1). Attempts to clone this additional *Eco* RI fragment have thus far been unsuccessful.

Pattern of Substitutions Among T15 V$_H$ Family Alleles. Fig. 6 presents an analysis of substitutions within the coding regions of the four gene segments of the B10.P T15 V$_H$ gene family compared with their BALB/c allelic counterparts. Certain general features of these results are apparent: (*a*) the number of substitutions differs greatly among the four family members compared: (*b*) substitutions are widely distributed but tend to exclude the hypervariable regions (only one of 31 replacement substitutions falls within a hypervariable region); (*c*) replacement substitutions account for 31 of 41 substitutions (76%), about what would be expected on a random basis, when all four gene segments are considered; however, all substitutions in the V3 pseudogene are of the replacement type (including one insertion); among the three functional V$_H$ gene segments there are 27 substitutions, of which 17 (63%) result in an amino acid replacement; and (*d*) the V1 gene segment of B10.P mice has accumulated only one-third as many
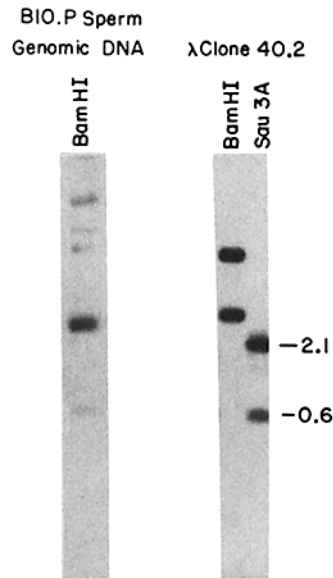
			HV1				HV2			
		10	20	30	40	50	60	70	80	90
BALB/c	T15	EVKLVESGGGLVQPGGSLRLSCATSGFTFSDFYMEWVRQPPGKRLEWIAASRNKANDYTTEYSASVKGRFIVSRDTSQSILYLQMNALRAEDTAIYYCARD
B10.P	VB2.4	...........S.R...............................A...G..
C57BL/10	293	...........S.R...............................A...G..
	1613	...........S.R...............................A...G..
	23169	...........S.R...............................A...G..
	2857	...........S.R...............................A...G..
	2313	...........S.R...............................A...G..

FIGURE 4. The λVB2.4 gene segment encodes PC-binding heavy chains from C57BL/6 mice. The amino acid sequence resulting from translation of the λVB2.4 nucleotide sequence is compared with amino acid sequences from PC-binding hybridomas previously determined by Clarke et al. (26) and with the sequence of the T15 heavy chain (11). The strain of origin of each sequence is noted at the left.

FIGURE 5.   The 2.4 and 4.4 kb genomic *Bam* HI fragments of the B10.P T15 V$_H$ family are linked. Shown are Southern blot patterns for B10.P sperm DNA and λ clone 40.2 DNA digested with *Bam* HI or *Sau* 3A (λ40.2 only) observed using the pS107V1 probe. The sizes of the individual bands (kb) are indicated.
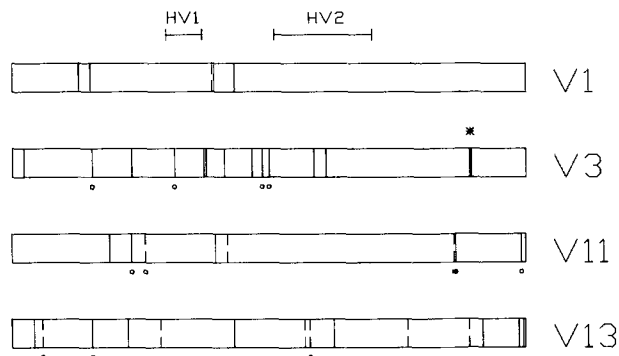
FIGURE 6.   Sequence substitutions in the T15 V$_H$ gene family. For each gene segment, a solid vertical line designates a replacement substitution distinguishing BALB/c and B10.P alleles; an interrupted line designates a silent substitution. The positions of conventionally defined hypervariable regions (37) are noted above the figure. An asterisk marks the site of a four nucleotide insertion in BALB/c relative to B10.P. Circles below the lines indicate substitutions that may possibly be attributable to gene conversion (see text for description).

substitutions as have the V3 and V13 gene segments (5, 14, 8, and 14 for V1, V3, V11, and V13, respectively).

Discussion

The T15 V$_H$ gene family in BALB/c mice includes three functional gene segments (V1, V11, and V13) and a pseudogene (V3) that are >88% homologous to each other (16). The V1 gene segment of this family encodes most V$_H$ regions

of PC-binding antibodies (11), and these antibodies are protective against infection with PC-bearing bacterial organisms (18). Thus, we suspect that this gene segment may be subject to considerable environmental selection pressure.

We have used the T15 V_H gene family to assess how V_H gene families change during evolution. In particular we sought to compare the divergence of the V1 element, perhaps subject to significant selection pressure, with that observed in the V3 pseudogene, which should drift in the complete absence of selection since it cannot be transcribed or translated (see below). Because of unique structural features of each of the members of the T15 V_H family, we succeeded in fully characterizing V_H alleles in two inbred mouse strains, thus enabling a detailed analysis of the diversification of this germline family.

T15 Gene Family Diverged Before the Divergence of Inbred Mouse Strains. Using a probe for the V1 gene, Southern blots revealed that most mice contain two to four similarly sized, hybridizing *Eco* RI or *Bam* HI restriction fragments at high stringency (Fig. 1). The T15 gene family, defined in this way, has undergone recent expansion and contraction events within murine species, since A/J mice appear to contain only two elements in the family while Lewis rats have at least 14 hybridizing bands. We chose to examine the V_H genes from mice of the Ig-1^b allotype since these likely separated from BALB/c at an early point in the divergence of inbred mice, perhaps one million years ago (28, 31). It should be noted that evidence derived from the study of mitochondrial DNA suggests that all inbred mice share a common ancestor distinct from typical wild mice (32). Thus, the time of divergence of inbred mice is uncertain.

B10.P DNA Contains Alleles for All Elements of the T15 Gene Family. Using DNA sequence analysis we identified the V1, V3, V11, and V13 alleles in the B10.P Ig-1^b T15 gene family (Figs. 2–5). In the case of V1, this analysis was straightforward, since only five nucleotide substitutions separate the BALB/c and B10.P alleles. Similarly, V3 could be easily identified by a characteristic deletion in the 3' rearrangement signals (Fig. 3), its immediate linkage to V1 (Fig. 5), and by characteristic leader and 5' flanking elements (Fig. 3 and see below). If we consider only the coding sequences, however, alleles within the T15 gene family cannot be easily distinguished. V11B10 is clearly most homologous to its BALB/c counterpart; however, V13B10 has diverged to such an extent that only consideration of flanking region substitutions enabled confident assignment of this allele. This rapid divergence of germline V_H sequences, similar to that observed by Loh et al. (15) in NP^b-related V_H genes, where the large number of substitutions precluded identification of alleles, reflects the high rate of nucleotide substitution in the T15 V_H family. We have analyzed the characteristics of these substitutions in some detail.

V1B10 Encodes Heavy Chains of PC-binding Antibodies in Ig-1^b Mice. V1B10 is functionally as well as structurally related to V1BALB since it participates in generating antibodies to PC. This is demonstrated in Fig. 4, which compares protein sequence data from C57BL/10 anti-PC antibodies with the translated gene sequence of λVB2.4. These protein sequences are identical with the germline λVB2.4 sequence, indicating that this gene segment is indeed the allele of V1BALB.

Evolution of the V3 Pseudogene. The sharp decline in homology between V3

and V13 within the leader segment (Fig. 3) defines what we believe to be the breakpoint in a recent gene duplication event that created V3 by homologous but unequal recombination. Since this duplication event would have destroyed the promoter region and the initiation codon of this gene segment, V3 likely was a pseudogene from the time of its creation. Subsequently, a deletion in the 3' rearrangement signals occurred, which is evident in both V3BALB and V3B10, and V3BALB accumulated two additional deleterious mutations after the time that the BALB/c and C57BL/10 chromosomes diverged. It is possible, though we believe less likely, that the deletion in the 3' rearrangement signals of V3 occurred first, rendering it a pseudogene, with subsequent scrambling of the 5' flanking sequences by a large deletion or insertion event. Resolution of these alternative formulations will require isolation of an allelic V$_H$ gene with an intact 5' or 3' flanking region.

If we presume that a common murine ancestor contained three members of the T15 V$_H$ gene family (analogous to V13, V11, and V1, and in that order on the chromosome), an unequal crossover event between V11 and a sequence 16 kb upstream of V1 would yield mice with four T15 V$_H$ family members, such as BALB/c and C57BL/10. The reciprocal product of this recombination would contain only V1 and V13 gene segments. This is the pattern seen in genomic blots using A/J DNA (Fig. 1). Examination of sequences upstream of the A/J V1 gene segment could therefore prove informative regarding this proposed gene duplication event.

Elements of the T15 Gene Family Evolve at Different Rates.   Summing over all four coding regions from the B10.P T15 gene family as compared with the analogous BALB/c T15 elements, there are 41 substitutions and one insertion in 1212 bp that were compared, for an overall divergence rate of 3.5%. In those genes which are expressed, the frequency of replacement substitutions is 17 of 27 or 63%, somewhat less than expected given the frequency of synonymous codons in the genetic code, which implies that selection is acting to conserve protein structure. Transitions account for 61% of all substitutions although, on a random basis, transversions should be twice as frequent. This observation suggests that mutations are occurring by a mechanism that is skewed from random.

Using the three-substitution model of Kimura (33), we have calculated the evolutionary rates at each codon position for each allele and for the γ2a and γ2b alleles of the same mouse strains (Table II; 34). Since BALB/c and C57BL/10 mouse strains diverged quite recently, these rates are equivalent to the number of substitutions divided by the number of codon comparisons (33). In comparing a variety of mammalian genes, substitution rates are typically highest for the third codon position (reflecting the high frequency of synonymous substitutions at this position) and lowest for the second codon position (33). As is visible by inspection, the V3 and V13 elements have fixed substitutions considerably more rapidly than the V1 gene segment. This is especially obvious when the silent substitution rate is considered (all third position substitutions in V1 and V13 are silent; Table II). Considerable evidence supports the use of the silent substitution rate as an evolutionary clock with a value of 5.1×10^{-9} substitutions per site per year (35). Using this estimate, the time of divergence of the V1 alleles corresponds

TABLE II
Evolutionary Rates in the T15 V_H Gene Family

	K1	K2	K3
V1	0.04 (0.02)	0.00	0.01 (0.01)
V3	0.04 (0.02)	0.07 (0.03)	0.03 (0.02)
V11	0.02 (0.01)	0.02 (0.01)	0.04 (0.02)
V13	0.02 (0.01)	0.06 (0.03)	0.06 (0.03)
γ2a*	0.07 (0.03)	0.09 (0.03)	0.07 (0.03)
γ2b*	0.01 (0.01)	0.003 (0.01)	0.01 (0.01)

K1, K2, and K3 are the rates at codon positions 1, 2, and 3, respectively
(33). The numbers in parentheses are standard errors.
*Calculated from sequences presented in reference 34.

to 0.5 million years, a value in accord with expectation (28, 31). Similar calculations for the V13 alleles yield divergence times of about three million years. It should be emphasized that since these rates are calculated using only silent site substitutions, the difference between the V1 and V13 evolutionary rates cannot be explained by selection at the level of mature protein. When the γ2a and γ2b constant region alleles from the BALB/c and C57BL/6 strains, encompassing 337 codons, are analyzed, a similar situation obtains (34). The γ2a alleles have fixed substitutions seven times more rapidly than the closely homologous and closely linked γ2b alleles.

Thus, the rates of evolution for these gene segments are very high; these rates vary even for linked elements within the same closely homologous V_H gene family; and the pattern of these substitutions is not completely consistent with differences resulting from selection pressure since even silent site substitution rates vary between different alleles.

T15 V_H Gene Family Substitutions in Detail.   The 13 substitutions in the V1 and V11 alleles are clustered and completely spare the hypervariable regions. Gap analysis (36) of the V3 and V13 alleles also suggests clustering, which perhaps implies that these substitutions arose in groups rather than as solitary events. In all, the hypervariable regions encompass 26 codons (31–37 and 52–71, numbering sequentially) or 25% of the total sequence compared (37). Of the 17 replacement substitutions in the V1, V11, and V13 alleles, only one falls within this region. The simplest interpretation of these data is that environmental selection has acted to preserve combining site sequences. This is the opposite result from that obtained by Loh et al. (15) where a predilection for substitutions in the hypervariable regions was observed. These data could be reconciled if comparisons in the NP[b] family, where specific alleles could not be identified, were made between nonallelic sequences that had been selected for binding site differences. Alternatively, selection may act differently in the NP[b] family, resulting in increased fixation of sequence differences within the combining site regions.

Gene Conversion in the T15 V_H Gene Family.   One attractive explanation for the rapid evolutionary rate of V_H genes invokes conversion events between homologous sequences (38–40). In examining the elements of the T15 V_H gene family, sequence substitutions that change a nucleotide present in the same position in

more than one family member to a different nucleotide, which is also represented in more than one family member, might easily be explained by conversion events. Considering only the V3, V11, and V13 sequences (which are >92% homologous among all six gene segments), it is unlikely that two alternative nucleotides would each be present in more than one family member at the same position. For any two alleles differing at one position, the probability of two alternative nucleotides at the same position among the remaining four sequences is $1 - (0.92)^4$ or ~0.28. Demanding that the substituted nucleotide be the correct substitution of three possible alternative nucleotides decreases the probability to (0.28×0.33) or ~9%. As shown in Fig. 6, 12 of 36 or 33% of the substitutions in V3, V11, and V13 fit this very restrictive definition of conversion. Excluded from this analysis are all potential conversion events that might have acted upon a nucleotide previously unique in the family. For example, BALB/c V3 differs from λVB4.4 at codon 24 by a G → T transversion that might reflect mutation of a progenitor G to a T in BALB/c, or a T → G conversion event with either V3 or V11 serving as substrate. Viewed in this way, the majority of the observed substitutions in the T15 V_H gene family might reflect a gene conversion process.

Exchange of sequence segments between nonallelic elements has been invoked to explain the pattern of substitutions in murine γ2a and γ2b genes (34) and the concerted evolution of immunoglobulin γ constant (C) region genes, in general, in mouse (39) and human C_α genes as well (40). Gene conversion might also explain the clustered pattern of substitutions in the T15 V_H gene family since these events have been demonstrated, for example, in the class I genes of the major histocompatibility complex, to involve short oligonucleotides (41). The highest rates of gene conversion would be anticipated among sequences with the highest levels of homology. In accord with this prediction, the V1 gene that is least homologous to the other family members has the lowest number of substitutions. In addition, the flanking region 5' of the V3 gene segment, which is <35% homologous with other family members and behaves as a single-copy element on genomic blots, is sequence identical between BALB/c and B10.P (R. M. Perlmutter, unpublished results).

A Comprehensive View of the Germline Antibody Repertoire.   By comparing the sequences of the four members of the T15 V_H gene family in B10.P mice with their allelic counterparts in BALB/c mice, certain inferences can be derived regarding evolution of antibody genes. First, despite the high rates of somatic mutation in antibody genes and the rapid evolutionary rates for both constant (34, 39) and variable (15) elements of these multigene families, there is considerable evidence for environmental selection of functional sequences. This is manifest in the relatively rapid rate of divergence at codon position two of the V3 pseudogene compared with other family members (Table II), and in the nearly complete sparing of hypervariable regions among the 17 replacement substitutions that have occurred in functional genes (Fig. 6). Secondly, the pattern of nucleotide substitutions among members of the T15 V_H gene family is quite nonrandom both with respect to type of substitution (transitions are favored) and with respect to placement. The percentage of these substitutions that could be explained by gene conversion events is much higher than expected, although

proof of gene conversion, in the sense of blocks of identifiable sequence clearly acquired from a demonstrated source, is lacking.

If gene conversion does indeed underlie many substitutions within the T15 V_H gene family, one paradoxical feature of these alterations and of similar conversion events described in class I genes of the H-2 complex (41) is that they act to increase heterogeneity within a gene family by shuffling regions of partial homology between groups of sequences with high overall homology. Gene conversion is proposed to explain the maintenance of homogeneous species-specific sequences in reiterated elements (e.g., ribosomal genes, 42). The mechanisms responsible for striking a balance between homogenization and diversification in multigene families are incompletely understood (43).

Lastly, we note that the high rate of substitution observed in the T15 V_H gene family is in accord with the extensive and apparently rapid divergence of inbred mouse strains tabulated by Fitch and Atchley (44). Although our data do not permit determination of the mechanism underlying this rapid divergence, the relatively high rate of synonymous substitutions within this family suggests that some genetic mechanism, perhaps gene conversion, acts to increase the number of substitutions per generation and that these substitutions are then acted upon by selection. The alternative explanation, that selection for heterozygosity has directed the inbreeding of mice (44), seems less likely, based on our data. Additional comparisons within the T15 V_H gene family should aid in interpreting the accelerated divergence of genes in inbred mice.

Summary

The T15 heavy chain variable region (V_H) gene family in BALB/c mice includes four elements each >88% homologous with the other. One of these elements, V1, encodes virtually all of the V_H regions in BALB/c antiphosphorylcholine antibodies, while another element, V3, is a pseudogene and cannot be transcribed or translated. We have examined the structural features of this V_H gene family in other mouse strains and, in particular, have cloned and sequenced the alleles of these gene segments present in B10.P mice. Each of the four B10.P sequences can be matched with its allelic counterpart in BALB/c mice. This represents the first successful analysis of allelism in antibody variable region gene segments. The V1B10.P allele, like its BALB/c counterpart, encodes most of the known phosphorylcholine binding heavy chains from C37BL/6 mice. Similarly, the V3B10.P gene segment is a pseudogene like V3BALB, although only two of four abnormalities present in the BALB/c allele are also present in the B10.P allele. Careful analysis of the specific substitutions observed in the T15 V_H gene family suggests that environmental selection for functional combining regions contributes significantly to the pattern of variation in the germline antibody repertoire. In addition, evidence is presented supporting frequent gene conversion events in the divergence of antibody genes.

References

1. Honjo, T. 1983. Immunoglobulin genes. *Annu. Rev. Immunol.* 1:499.

2. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature (Lond.).* 302:575.

3. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Goverman, T. Hunkapiller, M. B. Prystowsky, Y. Yoshikai, F. Fitch, T. W. Mak, and L. Hood. 1984. Mouse T cell antigen receptor: structure and organization of constant and joining gene segments encoding the β polypeptide. *Cell.* 37:1101.

4. Davis, M. M. 1985. Molecular genetics of the T-cell receptor β chain. *Annu. Rev. Immunol.* 3:537.

5. Chien, Y.-H, D. M. Becker, T. Lindsten, M. Okamura, D. I. Cohen, and M. M. Davis. 1984. A third type of murine T-cell receptor gene. *Nature (Lond.).* 312:31.

6. Kranz, D., H. Saito, M. Heller, Y. Takagaki, W. Haas, H. Eisen, and S. Tonegawa. 1985. Limited diversity of the rearranged T cell γ gene. *Nature (Lond.).* 313:752.

7. Brodeur, P. H., and R. Riblet. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. 100 Igh-V genes comprise 7 families of homologous genes. *Eur. J. Immunol.* 14:922.

8. Rocca-Serra, J., H. W. Matthes, M. Kaartinen, C. Milstein, J. Theze, and M. Fougereau. 1983. Analysis of antibody diversity: V-D-J mRNA nucleotide sequence of four anti-GAT monoclonal antibodies. A paucigene system using alternate D-J recombinations to generate functionally similar hypervariable regions. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:867.

9. Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore. 1981. Heavy chain variable region contribution to the NP family of antibodies: somatic mutation evident in a γ2a variable region. *Cell.* 24:625.

10. Perlmutter, R. M., J. L. Klotz, M. W. Bond, M. Nahm, J. M. Davie, and L. Hood. 1984. Multiple V$_H$ gene segments encode murine antistreptococcal antibodies. *J. Exp. Med.* 159:179.

11. Perlmutter, R. M., S. T. Crews, R. Douglas, G. Sorensen, N. Johnson, N. Nivera, P. J. Gearhart, and L. Hood. 1984. The generation of diversity in phosphorylcholine-binding antibodies. *Adv. Immunol.* 35:1.

12. Manser, T., S.-Y. Huang, and M. L. Gefter. 1984. Influence of clonal selection on the expression of immunoglobulin variable region genes. *Science (Wash. DC).* 226:1283.

13. Rodwell, J. D., P. J. Gearhart, and F. Karush. 1983. Restriction in IgM expression. IV. Affinity analysis of monoclonal anti-phosphorylcholine antibodies. *J. Immunol.* 130:313.

14. Weiss, S., K. Lehmann, W. C. Raschke, and M. Cohn. 1984. Mice completely suppressed for the expression of immunoglobulin κ light chain. *Proc. Natl. Acad. Sci. USA.* 81:211.

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

16. 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 antibody. *Cell.* 25:59.

17. Claflin, J. L., and S. Rudikoff. 1976. Uniformity in the clonal repertoire for the immune response to phosphorylcholine in mice: a case for a germline basis of antibody diversity. *Cold Spring Harbor Symp. Quant. Biol.* 41:725.

18. Briles, D. E., C. Forman, S. Hudak, and J. L. Claflin. 1982. Antiphosphorylcholine

antibodies of the T15 idiotype are optimally protective against *Streptococcus pneumoniae. J. Exp. Med.* 156:1177.

19. Huang, H., S. Crews, and L. Hood. 1981. An immunoglobulin V_H pseudogene. *J. Mol. Appl. Genet.* 1:93.

20. Blin, N., and D. W. Stafford. 1976. A general method for the isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* 3:2303.

21. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 284–285.

22. Loenen, W. A. M., and W. J. Brammar. 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. *Gene (Amst.).* 10:249.

23. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. *J. Mol. Biol.* 113:237.

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

25. Claflin, J. L., and J. M. Davie. 1974. Specific isolation and characterization of antibody directed to binding site antigenic determinants. *J. Immunol.* 114:70.

26. Clarke, S. H., J. L. Claflin, M. Potter, and S. Rudikoff. 1983. Polymorphisms in antiphosphorylcholine antibodies reflecting evolution of immunoglobulin families. *J. Exp. Med.* 157:98.

27. Lieberman, R. 1978. Genetics of the IgCH (allotype) locus in the mouse. *Springer Semin. Immunopathol.* 1:7.

28. Yonekawa, H., K. Moriwaki, O. Gotoh, J.-I. Hayashi, J. Watanabe, N. Mayashita, M. L. Petras, and Y. Tagashira. 1981. Evolutionary relationships among five subspecies of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA. *Genetics.* 98:801.

29. Mason, J. O., G. T. Williams, and M. S. Neuberger. 1985. Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. *Cell.* 41:479.

30. Parslow, T. G., D. L. Blair, W. J. Murphy, and D. K. Granner. 1984. Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. USA.* 81:2650.

31. Ferris, S. D., R. D. Sage, E. M. Prager, U. Ritte, and A. C. Wilson. 1983. Mitochondrial DNA evolution in mice. *Genetics.* 105:681.

32. Ferris, S. D., R. D. Sage, and A. C. Wilson. 1982. Evidence from mtDNA sequences that common laboratory mice are descended from a single female. *Nature (Lond.).* 295:163.

33. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* 78:454.

34. Ollo, R., and F. Rougeon. 1983. Gene conversion and polymorphism: generation of mouse γ2a chain alleles by differential gene conversion by γ2b chain gene. *Cell.* 32:515.

35. Miyata, T., T. Yasunaga, and T. Nishida. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proc. Natl. Acad. Sci. USA.* 77:7328.

36. Gearhart, P. J., and D. F. Bogenhagen. 1983. Clusters of point mutations are found exclusively around rearranged antibody variable region genes. *Proc. Natl. Acad. Sci. USA.* 80:3439.

37. Kabat, E. A., T. T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. 1983. Sequences of Proteins of Immunological Interest. United States Public Health Service, National Institutes of Health, Bethesda, MD. p. 129.

38. Baltimore, D. 1981. Gene conversion: some implications for immunoglobulin genes. *Cell.* 24:592.
39. Hayashida, H., T. Miyata, Y. Yamawaki-Kataoka, T. Honjo, J. Wels, and F. Blattner. 1984. Concerted evolution of the mouse immunoglobulin gamma chain genes. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:2047.
40. Flanagan, J. G., M.-P. Lefranc, and T. H. Rabbits. 1984. Mechanisms of divergence and convergence of the human immunoglobulin $\alpha 1$ and $\alpha 2$ constant region gene sequences. *Cell.* 36:681.
41. Pease, L. R., D. H. Schulze, G. M. Pfaffenbach, and S. G. Nathenson. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 80:242.
42. Nagylaki, T., and T. D. Petes. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics.* 100:315.
43. Hood, L., J. H. Campbell, and S. C. R. Elgin. 1975. The organization, expression and evolution of antibody genes and other multigene families. *Annu. Rev. Genet.* 9:305.
44. Fitch, W. M., and W. R. Atchley. 1985. Evolution in inbred strains of mice appears rapid. *Science (Wash. DC.).* 228:1169.