Allelic immunoglobulin V_H genes in two mouse strains: possible germline gene recombination

Justus B.Cohen and David Givol*

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot, Israel

Communicated by D.Givol Received on 13 July 1983; revised on 19 August 1983

The nucleotide sequence of two germline immunoglobulin heavy chain variable region (V_H) genes of mouse BALB/c origin was determined. These two genes are highly homologous to each other. They both have the unusual codon CCT for proline at position 7, which so far has been found only in a specific set of V_H genes, called the NP^b family. We show that the two V_H genes belong to this set. One of our BALB/c genes, V_H 124, is more homologous to a C57BL/6 NP^b V_H gene than to any BALB/c V_H gene, and we propose that these two genes are alleles. A comparison of the substitutions between these two genes with published sequences of all other BALB/c and C57BL/6 NP^b V_H genes reveals evidence for past homologous recombination events between related germline V_H genes. Homologous recombination may play an important role in the diversification of germline immunoglobulin V_H genes.

Key words: V_H gene/germline/gene recombination

Introduction

C57BL/6 mice generate a closely related set of antibodies in response to immunization with conjugates of the hapten NP [(4-hydroxy-3-nitrophenyl)acetyl-]. These antibodies can be distinguished from all other antibodies on the basis of a specific antigenicity of the variable portion of the molecules. This antigenicity (idiotype) has been termed NP^b idiotype. Since all C57BL/6 anti-NP antibodies have λ light chains (Jack *et al.*, 1977), for which only two germline V_{λ} genes exist in the mouse (Tonegawa et al., 1978), work on NP^b antibodies has mostly focused on characterizing the variable regions of the heavy chains (V_H). The V_H regions of these antibodies are encoded by a highly homologous set of germline V_H genes in the C57BL/6 mouse, and it is likely that most of them are encoded by a single member of this set called V_H186.2 (Bothwell et al., 1981). Restricted variability among the V_{H} regions of different NP^{b} antibodies is the result of somatic mutations generated during B cell maturation, and the use of different D (and possibly J_{H}) gene segments. One of the consistent features that distinguish the C57BL/6 NP^b-related V_H gene set from all other V_H genes is the presence of a codon (CCT) for proline at position +7(Bothwell et al., 1981).

BALB/c mice produce an anti-NP antibody response, which is similarly homogeneous with respect to idiotype (Karjalainen, 1980; White-Scharf and Imanishi-Kari, 1982). Although BALB/c anti-NP antibodies share certain antigenbinding characteristics with their C57BL/6 counterparts (Karjalainen, 1980; White-Scharf and Imanishi-Kari, 1981; 1982), they lack the NP^b idiotype and their idiotype is called NP^a,

*To whom reprint requests should be sent.

suggesting that perhaps C57BL/6 $NP^b\ V_H$ regions and BALB/c NP^a V_H regions are encoded by an allelic set of genes (Karjalainen, 1980). Recently, however, Loh et al. (1983) showed that the nucleotide sequence of the mRNA coding for a typical BALB/c NPa V_H region greatly differs from the NP^b-related V_H gene set, so that it is likely that BALB/c mice use different germline V_H genes in their response to NP. In the same study Loh et al. also showed that BALB/c mice do carry a set of genes, which is closely related to the C57BL/6 NP^b V_H gene set. This BALB/c set of V_H genes was called 'NPb-equivalent'. The use of different genes in the anti-NP response by BALB/c mice is therefore not due to the absence of V_H genes which are similar to those used by C57BL/6 mice. In fact, BALB/c mice are capable of producing immunoglobulins, which carry the NPb idiotype of C57BL/6 anti-NP antibodies (Takemori et al., 1982). These BALB/c NP^b immunoglobulins, which can be induced by immunization with antibodies that are specifically directed against the NP^b idiotype, however, do not bind NP. It is thus possible that their V_H regions are encoded by the BALB/c NP^b-equivalent V_H genes and that due to some minor differences between the two sets of genes the antibody formed does not bind NP.

The comparison by Loh *et al.* (1983) of the highly homologous NP^b-related (C57BL/6) and NP^b-equivalent (BALB/c) sets of V_H genes is relevant to our understanding of the molecular basis of antibody specificity and idiotypes. Their study also provides direct insight into the evolution of Ig V_H genes, and the mechanisms that operate to conserve or diversify germline V_H gene sequences. One of the conclusions from their work is, that it is very difficult to pinpoint allelic V_H genes in the NP^b sets of V_H genes from BALB/c and C57BL/6 mice. They attribute this finding to a frequent occurrence of gene conversion between V_H genes which obscures allelic relationships. However, it is difficult to draw definite conclusions from their comparisons, because some members of each set may not have been identified.

Here we present evidence, based on nucleotide sequence determination, for the presence of two additional members of the NP^b-equivalent V_H gene set in the BALB/c germline genome. Whereas at least three of the five NPb-equivalent genes of Loh et al. (1983) are pseudogenes, these two new genes seem to be normal, intact V_H genes. One of our genes, $V_{\rm H}$ 124, is 97.1% homologous with one of the C57BL/6 genes, V_H102. This pair of genes represents the most closely related pair of all analysed NP^b V_H genes between the two strains. Our data indicate that $V_H 124$ and $V_H 102$ may be allelic V_H genes. Interestingly, these genes diverged rather extensively in the sequence of their intron, while the intron sequence of $V_H 102$ is identical to that of two other BALB/c genes. Comparison with the intron sequences of other BALB/c and C57BL/6 genes suggests that the C57BL/6 allele of BALB/c gene V_H124 has acquired sequence information from other C57BL/6 genes after the separation of the two mouse strains to generate the present day $V_{\rm H}102$.



Fig. 1. Strategy used for nucleotide sequence determination of $V_H 104B$ (A) and $V_H 124$ (B). Black bars represent coding regions for the leader (L) and V_H gene segment. M + G: sequence determined by the chemical degradation procedure (Maxam and Gilbert, 1980); S: sequence determined by the dideoxy chain-termination method (Sanger *et al.*, 1980). (A), Numbers above the coding blocks refer to amino acid positions. (B), Hatched bars: pBR322 vector sequences; wavy lines: M13 vector sequences. The 5' *Eco*RI-*PstI* fragment of plasmid p124 (upper line) was subcloned into M13 vector mp8, and the 3' *PstI-Hind*III fragment into M13 vector mp9 (Messing and Vieira, 1982), and the V_H coding sequences determined by the dideoxy chain-termination procedure. The 3' *PstI-Hind*III fragment present in vector mp9 was recloned into pBR322 as a *Bam*HI-*Hind*III fragment, using the *Bam*HI site in the polylinker region of vector mp9 (Messing and Vieira, 1982) as a linker.

Results

Isolation and sequencing strategy of immunoglobulin V_H genes

We have described the isolation of four Charon 4A recombinant clones, which contain mouse BALB/c germline V_H genes, and we reported the nucleotide sequences of five of the six genes present in these clones (Givol *et al.*, 1981; Cohen *et al.*, 1982). Here we focus on the sixth gene, V_H 104B, which is physically linked to V_H 104A (previously called pCh104, Givol *et al.*, 1981), and on a related germline gene, V_H 124, which was isolated from another Charon 4A genomic DNA clone. This clone (Ch124) also contains a diverged V_H pseudogene (Cohen and Givol, 1983).

Gene V_H104B is located on a ~7.8-kb *Eco*RI fragment in the genomic clone Ch104 (Givol *et al.*, 1981). We isolated this fragment, sonicated it and used the collection of fragments in the size range of ~300-700 bp for shotgun cloning into M13 bacteriophage vector mp8 (Deininger, 1983). Clones that hybridized to our V_H probe were used for determination of the V_H104B nucleotide sequence by the dideoxy chain termination method of Sanger *et al.* (1980), as shown in Figure 1A.

The construction of Ch124 subclone p124 is described by us elsewhere (Cohen and Givol, 1983). This subclone contains a 1.65-kb *Eco*RI-*Hind*III fragment carrying gene V_H124 . Figure 1B shows the strategy used to determine the nucleotide sequence of this gene using the procedures of Maxam and Gilbert (1980) and Sanger *et al.* (1980).

V_H 124 and V_H 104B are homologous to the NP^b V_H genes of C57BL/6

Figure 2 shows the nucleotide sequence of $V_H 124$ and

V_H104B. Both genes have the structure of a typical mouse germline V_H gene: a V_H gene segment of 98 codons, which is linked to the last four codons of the hydrophobic signal peptide of V_H, and separated from the leader segment by an intervening sequence of 82-84 bp. The V_H gene segment is followed by the signals for V_H-D recombination: the conserved heptamer CACAGTG, which is separated by 23 bp from the less conserved nonamer TAGAAACCC in V_H124, like in numerous other germline V_H genes (Early *et al.*, 1980; Bothwell *et al.*, 1981; Crews *et al.*, 1981; Givol *et al.*, 1981).

Genes V_H124 and V_H104B are more homologous to each other in the V_H gene segment (92.5%) than any of the five genes of the MPC11 V_H gene family that we previously compared (of which the physically linked V_H108A and V_H108B were the most homologous pair, 91.1%) (Givol et al., 1981; Cohen *et al.*, 1982). We have shown before that these five V_{H} genes differ much more in their complementarity determining regions (CDR) than in their framework regions (FR) (Givol et al., 1981). The same holds true for $V_H 124$ and $V_H 104B$ in CDRII - 19.6% substitutions versus 5.9, 4.8 and 5.2%, respectively, in the FRs - but CDRI is identical between the two genes. V_H124 and V_H104B differ from our previously analysed V_H genes in two other respects: (i) both have only a single PstI site in their coding region, whereas the previously analysed genes all have at least two PstI sites, and (ii) both have the unusual codon CCT for proline at position +7 of the V_H coding region. These three findings urged us to compare the sequences of V_H124 and V_H104B with those of the seven germline V_H genes of mouse strain C57BL/6 origin, which have been described by Bothwell et al. (1981), and which have the same three unusual features.

Figure 2 includes the nucleotide sequences of two C57BL/6 V_H genes as reported by Bothwell *et al.* (1981). Gene V_H 186.2 was shown by these authors to code for the V_H regions in a majority of (or possibly all) NP binding antibodies in C57BL/6 mice. Such antibodies have a common antigenicity, which is determined by their V regions (idiotype), and which is called NP^b idiotype. The family of C57BL/6 V_H genes, which is homologous to V_H 186.2, was therefore called the NP^b-related V_H gene set (Bothwell *et al.*, 1981; see also Introduction). C57BL/6 gene V_H 102, which is also given in Figure 2, belongs to this V_H gene set (Bothwell *et al.*, 1981).

The comparison in Figure 2 of mouse BALB/c genes $V_H 124$ and $V_H 104B$ with C57BL/6 genes $V_H 186.2$ and $V_H 102$ shows that the BALB/c and the C57BL/6 genes are highly homologous. The nucleotide sequence of $V_H 104B$ shows 92.5% identity in its V_H gene segment with that of V_H 186.2 (compare with 92.5% for $V_H 104B/V_H 124$) and 91.8% with $V_H 102$. The homology of $V_H 124$ with the two C57BL/6 genes is even higher: 94.1% with $V_H 186.2$ and 97.1% with $V_H 102$. These data unambiguously establish that $V_H 104B$ and $V_H 124$ belong to a BALB/c V_H gene set, which is analogous to the C57BL/6 NP^b-related V_H gene set.

V_H 124 shares extensive stretches of sequence with two C57BL/6 V_H genes

A more detailed analysis of the sequence comparison of Figure 2 reveals an interesting pattern of substitutions between V_H124 and the two C57BL/6 genes. Starting just upstream to the 3' intron/exon border, V_H124 is identical in nucleotide sequence to $V_H186.2$ over a stretch of 152 bp, including the first 42 codons of the V_H gene segment. The other C57BL/6 gene, V_H102 , differs at three positions from V_H124 in this region. Going further towards the 3' end of the genes,

-10 -19 Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ser Thr Ala Thr G ATG GGA TGG AGC TGT ATC ATC CTC TTC TTG GTA TCA ACA GCT ACA G/GTAAGGGGGC TCACAGGTAA BALB/c V_µ124 Ala --- --- G--C57BL/6 V_H102 --- --- -/----- ----Met Ala Ala C57BL/6 V_H186-2 -----G ---- -C- G------ --- -/-----Ser Ala BALB/c V_H104B --- --- --- ---G-- --- -/---- -----Cly Val His V_H124 GCAGGCTTGA GAACTGGCCA TACCTGTGGG TGAAAATGAC ATCCACTCTC TCTTTCTCTC CACAG/GT GTC CAC V_H102 -----T-G C-----T-G -----V_H186-2 V_H104B +10 Ser Gln Val Gln Leu Gln Gln Pro Gly Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys TCC CAG GTC CAA CTG CAG CAG CCT GGC GCT GAG CTT GTG AAG CCT GGG GCT TCA GTG AAG V_H124 V_H102 V_H186-2 V_H104B +20 +30 CDRI Leu Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser Tyr Trp Met His Trp Val Lys Gln CTG TCC TGC AAG GCT TCT GGC TAC ACC TTC ACC AGC TAC TGG ATG CAC TGG GTG AAG CAG V_H124 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---V_H102 V_H186-2 A1a V_H104B -C- --- ---+40 Arg Pro Gly Gln Gly Leu Glu Trp Ile Gly Clu Ile Asp Pro Ser Asp Ser Tyr Thr Asn AGG CCT GGA CAA GGC CTT GAG TGG ATC GGA GAG ATT GAT CCT TCT GAT AGT TAT ACT AAC V_H124 Arg Asp His --- --- C --- --- G-- --- --- AG- --- C-- --- --- G-- --- G--V_H102
 Arg
 Arg
 Asn Ser Gly Gly
 Lys

 ------G ------G
 -------G
 -------G
 -------G
Asn Ser Gly Gly Lys V_H186-2 His Asn Cys Gly Asn Ile V_H104B --- --- AA- TG- G-- A-- -T --- C-- --- AA- TG- G-- A-- -T- ---+70 +60 Tyr Asn Gln Lys Phe Lys Gly Lys Ala Thr Leu Thr Val Asp Lys Ser Ser Ser Thr Ala TAC AAT CAA AAG TTC AAG GGC AAG GCC ACA TTG ACT GTA GAC AAA TCC TCC AGC ACA GCC V_H124 --- --- --- --- --- --- --- --- --- --- --- ---V_H102
 Glu
 Ser
 Leu
 Pro

 --- G-G
 --- C --- --- Glu
 Thr

 Glu
 Leu
 Thr
 C --- --- Glu
 Thr
V_H186-2 ---- --- --- --- C-- --- -C- ---- -C- ----V_µ104B --- --- G-G +90 +98 +80 Tyr Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys Ala Arg TAC ATG CAG CTC AGC AGC CTG ACA TCT GAG GAC TCT GCG GTC TAT TAC TGT GCA AGA V_H124 Ile --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---V_H102 V_H186-2 Val Asp V₁₁104B V_H124 CACAGTOTTG TAACCACATT CTGAGAGTGT TAGAAACCCT GAGGAGTAGT AAACTGTCCT GAGACT V_H102 V_H186-2 V_H104B

Allelic immunoglobulin V_H genes in two mouse strains

.

Fig. 2. Nucleotide sequence of BALB/c genes $V_H 124$ and $V_H 104B$, in comparison with C57BL/6 NP^b genes $V_H 186.2$ and $V_H 102$ from Bothwell *et al.* (1981). Dashes indicate identity with the V_H124 sequence. Open spaces stand for nucleotide deletions, and insertions are indicated by a letter above the line of the nucleotides. Amino acid residues predicted from the nucleotide sequence are given above the line of the codons. The $V_H 124$ intron sequence is given in italics. Complementarity determining regions (CDRI and II) are overlined, and the 3' signals for V_H -D joining are boxed.

| Twole It Companion of percentage nonitology contraction of H Henric | Table I. | Comparison | of percentag | e homology | between NP ^I | V _H genes ^{a,b} |
|---|----------|------------|--------------|------------|-------------------------|-------------------------------------|
|---|----------|------------|--------------|------------|-------------------------|-------------------------------------|

| A. Range of homology | | | | |
|---|---|---|-------------------------------------|--------------------------------------|
| BALB/c V _H 124 vs. all BALB/c genes | C57BL/6 V _H 102 vs. all C57BL/6 genes | BALB/c V _H 124 vs. C57BL/6 V _H 102 | All BALB/c NP ^b genes | All C57BL/6 NP ^b genes |
| 87.2-92.5% | 92.1-94.4% | 97.1% | 85.6-94.4% | 90.2-99.0% |

B. Most homologous pairs of NP^b V_H genes within and between strains

| BALB/c | | C57BL/6 | | BALB/c-C57BL/6 | | |
|--|-------|---|-------|---------------------------------------|-------|--|
| V _H 33-V _H 28 | 94.4% | V _H 186.2-V _H 145 | 99.0% | V _H 124-V _H 102 | 97.1% | |
| V _H 124-V _H 104B | 92.5% | V _H 186.2-V _H 186.1 | 98.7% | V _H 3-V _H 186.2 | 95.7% | |
| V _H 124-V _H 33 | 92.5% | V _H 186.1-V _H 145 | 97.7% | V _H 33-V _H 23 | 94.4% | |
| $V_H^3 - V_H^3$ | 92.5% | V _H 186.2-V _H 6 | 95.7% | V _H 28-V _H 6 | 94.4% | |

^aBALB/c $V_{\rm H}$ gene sequences used in the comparison are from Loh *et al.* (1983) and this paper, and C57BL/6 $V_{\rm H}$ gene sequences are from Bothwell *et al.* (1981).

^bSequences compared include only the V_H gene segment (codons 4-98).

Table II. Analysis of substitutions between $V_H 124$ and $V_H 102$ by comparison with nucleotides present in these positions in other NP^b V_H genes^a

| | | Substitutions at codon number | | | | | | | |
|-----------------------------|----------------------|-------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|-------------|-------------|
| | | 20 | 42 | 48 | 52 | 57 | 98 | 11 | 50 |
| BALB/c | V _H 124 | CTG | GGA | ATC | GAT | TAT | AGA | CTT | <u>GA</u> G |
| C57BL/6 | V _H 102 | GTG | GGC | ATT | CAT | GAT | ATA | СТ <u>G</u> | <u>AG</u> G |
| BALB/c | V _H 104B | С | Α | Т | С | Α | G | G | GA |
| | V _H 3 | Α | Α | Т | Т | G | Т | Т | AG |
| | V _H 33 | Т | Α | Т | Α | G | Т | G | GA |
| | V _H 28 | Т | Α | Т | Α | Α | G | G | AA |
| | V _H 5 | Α | Α | Т | Α | Т | G | G | TA |
| | V _H 31 | Α | Α | Т | Т | Α | G | G | TG |
| C57BL6 | V _H 186.2 | С | Α | Т | G | G | G | Т | AG |
| | V _H 186.1 | С | Α | Т | G | G | G | Т | AG |
| | V _H 145 | С | Α | Т | G | G | G | G | AG |
| | V _H 23 | С | Α | Т | Α | G | G | Т | AA |
| | V _H 6 | С | Α | Т | Т | Α | G | G | AA |
| | V _H 3 | С | Α | Т | Т | G | G | G | AA |
| Suggested point mutation in | | V _H 102 | V _H 102 | V _H 124 | V _H 102 | V _H 124 | V _H 102 | ? | ?? |

^aBALB/c V_H gene sequences were from Loh *et al.* (1983) and this paper (V_H 124 and V_H 104B) and C57BL/6 V_H gene sequences were from Bothwell *et al.* (1981).

we find the opposite pattern of identities: between codons 58 and 98, $V_H 124$ differs in seven positions from $V_H 186.2$, whereas its sequence is identical to that of $V_H 102$ in this region, except for a single base substitution in codon 98. The most straight-forward explanation to account for this alternating pattern of identities of $V_H 124$ with two different V_H genes is, that $V_H 124$ has exchanged sequences in the past with both $V_H 186.2$ and $V_H 102$. However, a detailed comparison with all C57BL/6 and BALB/c NP^b genes suggests a different explanation.

BALB/c V_H124 and C57BL/6 V_H102 may be allelic genes

Since $V_H 124$ originates from a different mouse strain than $V_H 186.2$ and $V_H 102$, the question arises how the latter two genes could have interacted with the first one to exchange sequences. In an attempt to answer this question we compared BALB/c gene $V_H 124$ with the sequences of five BALB/c genes, which are highly homologous to the C57/BL6 NP^b-related set of V_H genes, and which were described recently by Loh *et al.* (1983; see Introduction). This comparison, given in Table I, shows that $V_H 124$ is considerably less homologous to

the BALB/c NP^b-equivalent V_H genes (including V_H 104B) than to C57BL/6 gene $V_{\rm H}$ 102. No extensive regions of identity between V_H124 and any of the other BALB/c genes were found (data not shown). When we analysed the C57BL/6 NP^b-related V_H genes of Bothwell *et al.* (1981), we found a similar phenomenon for gene $V_H 102$: this gene is much less homologous to the NP^b-related V_H genes of its own strain than to BALB/c gene V_H124 (Table I). In addition, BALB/c $V_{\rm H}$ 124 and C57BL/6 $V_{\rm H}$ 102 are the most homologous pair of NP^b genes between the two strains, as shown in Table I. We conclude from these data that BALB/c gene V_H124 and C57BL/6 gene $V_{\rm H}$ 102 could in fact be allelic genes. The small number of substitutions in the V_H gene segment (2.9%) and leader sequence (2.2%) between these genes could then be the result of evolutionary divergence since the separation of the two mouse strains. Indeed, analysis of the sites of substitution between $V_H 102$ and $V_H 124$ in the other BALB/c and C57BL/6 NP^b genes shows that at least six of the nine substitutions in the V_H gene segment are probably simple point mutations. Table II lists these substitutions. One very clear example from this table is the substitution in codon 42



Fig. 3. Schematic representation of the substitutions between the allelic genes V_H102 (C57BL/6; Bothwell *et al.*, 1981) and V_H124 (BALB/c; this work), and comparison with the pattern of substitutions between V_H102 and BALB/c genes V_H3 and V_H33 from Loh *et al.* (1983). Open boxes indicate coding regions for the leader (L) and V_H gene segments. Base substitutions relative to the V_H102 sequence are indicated by vertical lines. Codon numbers are given at the top of the figure.

between V_H124 (GGA) and V_H102 (GGC). All other BALB/c and C57BL/6 genes have an A in the last position of this codon, so that it is evident that the C in the V_H102 codon is the result of point mutation after the separation of the two mouse strains. Table II shows that four of the point mutations occurred in the V_H102 sequence and two in the V_H124 sequence. This is remarkable, because the C47BL/6 NP^b V_H gene family internally diverged much less than its BALB/c counterpart, as can be concluded from the data in Table I.

Three substitutions between $V_H 124$ and $V_H 102$ cannot be easily explained as resulting from point mutation (Table II). Codon 11 is CTT in $V_H 124$ and CTG in $V_H 102$. As shown in Table II, the other BALB/c and C57BL/6 genes have either T or G at the position of substitution. Because the CTT/CTG substitution does not result in an amino acid substitution, it cannot be due to selection on the protein level. Hence, if $V_H 124$ and $V_H 102$ are allelic V_H genes, some other mechanism than the fixation of point mutations must have generated this substitution. The two substitutions in codon 50 may also be due to such a mechanism.

Intron sequence divergence seems to result from gene recombination

Intron sequences in general diverge faster than coding regions, so that it may not be surprising that, in contrast to their coding regions, $V_H 124$ and $V_H 102$ differ considerably (14.3%) in the sequence of their intron. We noticed however that two of the BALB/c NP^b-equivalent V_H genes of Loh et al. (1983), V_{H3} and V_{H33} , have intron sequences that are identical to that of V_H102. In their coding regions on the other hand, these two genes differ rather extensively from V_H102 (7.5% each), as shown in Figure 3. Hence different segments of the same gene diverged to a different extent from $V_{H}102$ not only in the case of $V_{H}124$, but also in the case of V_{H3} and V_{H33} . This finding seems unusual for diverging genes, unless some other mechanism has contributed to their divergence in addition to the fixation of point mutations. Homologous recombination, which may exchange segments between related genes, either in a reciprocal or a nonreciprocal fashion, could be such a mechanism. It could be responsible for the observed phenomenon, and thus contribute to the diversification of germline $V_{\rm H}$ genes.

Discussion

Here we describe the nucleotide sequences of two mouse BALB/c germline V_H genes. These two genes differ from the five that we previously described in three aspects: they are identical in CDRI, they have a single *PstI* site in their coding region, and they have a codon for proline at position 7 of the V_H segment. We have shown that these genes belong to a set of V_H genes, described recently by Loh *et al.* (1983), which is equivalent to the NP^b-related set of V_H genes present in C57BL/6 mice.

One of the two BALB/c genes, $V_H 124$, and the C57BL/6 gene $V_H 102$ are more closely related to each other in their V_H gene segment than to any gene of their own strain. We analysed the substitutions between these two genes and conclude that they may be allelic V_H genes, which have diverged mainly by fixing point mutations. Loh *et al.* (1983) have shown that matching of allelic pairs of V_H genes is not a simple matter, even if one compares strains of mice that may have separated as recently as 0.3 million years ago. It is evident from their data that allelic NP^b V_H genes must have diverged considerably in this short time span. The divergence between the V_H segments of $V_H 102$ and $V_H 124$ (2.9%) does not therefore seem to be unusually high for allelic V_H genes.

The intron sequence of $V_H 102$ differs extensively from that of $V_{\rm H}124$, but is identical to those of two other BALB/c NP^b-equivalent genes, V_H3 and V_H33 . A past exchange of sequences between related BALB/c V_H genes may be responsible for this phenomenon. Since the V_H102 intron sequence is however also identical to that of two genes of its own strain - V_H145 (except for a single point mutation) and V_H23 (Bothwell *et al.*, 1981) – we find it more likely that $V_{\rm H}102$ acquired its present intron sequence from one of these two $C57BL/6 V_{H}$ genes by gene conversion after the separation of the two mouse strains. A major reason for this suggestion is that none of the seven C57BL/6 NP^b V_H genes of Bothwell et al. (1981) has an intron sequence that could possibly be allelic to that of V_H 124. Once it is accepted that recombination in the form of gene conversion may occur between homologous germline V_H genes, it becomes apparent that this may also be the mechanism that generated the three substitutions in the coding regions of V_H124 and V_H102 that are not the result of point mutation. In addition, it may explain the restricted divergence ($\sim 1 - 10\%$) among the C57BL/6 NP^b genes relative to the more extensive divergence ($\sim 5-15\%$) among the allelic BALB/c NP^b V_H genes (see Table I). Gene conversion may thus both correct genes against each other (e.g., the C57BL/6 genes), and diversify them (e.g., the substitutions in codons 11 and 50 in V_H124/V_H102).

Work from other laboratories has lately stressed the potential of gene conversion to diversify Ig variable region genes. Bentley and Rabbitts (1983) showed most clearly the effect of gene conversion on two recently duplicated germline V_x genes. Gene conversion has also been proposed as the mechanism that altered the expressed V_H region in a somatic variant of an anti-NP hybridoma (Dildrop *et al.*, 1982). With respect to germline V_H genes it has proved to be extremely difficult to find clear examples of gene conversion because of the appreciable rate at which point mutations accumulate in these genes (Clarke *et al.*, 1982; Loh *et al.*, 1983). Our analysis of BALB/c gene V_H 124 enables us to pinpoint a gene

J.B.Cohen and D.Givol

conversion event between its C57BL/6 allele $V_H 102$ and another C57BL/6 gene only because the genes of this particular V_H gene family are still relatively homogeneous in sequence. Studies on allelic genes from other conserved Ig V_H gene families, such as the T15 family (Crews *et al.*, 1981; Rudikoff and Potter, 1980), may thus yield further information on the role of gene conversion in the evolution of germline Ig V_H genes. For a discussion of gene conversion in C_H alleles see also Ollo and Rougeon (1983).

After this work was submitted we obtained from K.Rajewsky the amino acids sequence of the V_H region of antibody 260/2 (Dildrop,R., Bovens,H., Siekevitz,M., Beyrewther,K. and Rajewsky,K., in preparation). This antibody is produced by a hybridoma derived from cells of C57BL/6 mice immunized with anti-idiotope antibody Ac 38 (Roth *et al.*, 1979; Takemori *et al.*, 1982). It contains the N_P^b idiotope but does not bind N_P ligands. So far the V_H region of 260/2 has been sequenced between amino acids 13 and 81. This sequence shows only one difference (Met instead of Lys at position 13) with the predicted protein sequence of V_H124 described here, which is less than the four differences (three of them in CDR2), between V_H124 and V_H102 in this region (Figure 2).

These new data suggest that C57BL/6 mice may have another gene, which is more similar to $V_{\rm H}124$ than $V_{\rm H}102$. However, no firm conclusions can be drawn before such a gene has been identified and its DNA sequence determined. The characterization of the V_H gene coding for antibody 260/2 will be of great interest, since this antibody does not bind N_P and may provide a clearer example of V_H gene allelism between mouse strains. In addition, the sequence of its intron may yield more information concerning the evolutionary history of BALB/c and C57BL/6 MP^b V_H genes. If this putative V_H gene is more homologous to V_H 124 than $V_H 102$, the conservation of allelic V_H genes would be even more impressive in comparison with the 10% divergence of allelic C_H genes between the same two mouse strains (Schreier et al., 1981; Ollo and Rougeon, 1983). On an evolutionary scale it seems that variable region genes are relatively constant, and constant region genes are relatively variable.

Materials and methods

The isolation of germline V_H gene containing Charon 4A recombinant clones had been described (Givol *et al.*, 1981). Subcloning of appropriate restriction enzyme fragments into plasmid vector pBR322 and bacteriophage M13 vectors mp8 and mp9 (Messing and Vieira, 1982) is described elsewhere (Cohen and Givol, 1983). Shotgun cloning of sonicated DNA into M13 vector mp8 was performed according to Deininger (1983), and has been detailed by us before (Rechavi *et al.*, 1983). DNA nucleotide sequences were determined by the chemical degradation method (Maxam and Gilbert, 1980), as previous ly described (Cohen *et al.*, 1982), and by the dideoxy chain-termination procedure according to Sanger *et al.* (1980).

Acknowledgements

We thank Dr.S.Fields for assistance in M13 cloning and sequencing, and Dr.K.Rajewsky for communicating to us the sequence of antibody 260/2 prior to publication.

References

- Bentley, D.L. and Rabbitts, T.H. (1983) Cell, 32, 181-189.
- Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) Cell, 24, 625-637.
- Clarke, S.H., Claflin, J.L. and Rudikoff, S. (1982) Proc. Natl. Acad. Sci. USA, 79, 3280-3284.
- Cohen, J.B and Givol, D. (1983) EMBO J., 2, 1795-1800.

- Cohen, J.B., Effron, K., Rechavi, G., Ben-Neriah, Y., Zakut, R. and Givol, D. (1982) Nucleic Acids Res., 10, 3353-3370.
- Crews, S., Griffin, J., Huang, H., Calame, K. and Hood, L. (1981) Cell, 25, 59-66.
- Deininger, P.L. (1983) Anal. Biochem., 129, 216-223.
- Dildrop, R., Bruggemann, M., Radbruch, A., Rajewsky, K. and Beyreuther, K. (1982) *EMBO J.*, 1, 635-640.
- Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell, 19, 981-992.
- Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. and Cohen, J.B. (1981) Nature, 292, 426-430.
- Jack, R.S., Imanishi-Kari, T. and Rajewsky, K. (1977) Eur. J. Immunol., 7, 559-565.
- Karjalainen, K. (1980) Eur. J. Immunol., 10, 132-139.
- Loh, D.Y., Bothwell, A.L.M., White-Scharf, M.E., Imanishi-Kari, T. and Baltimore, D. (1983) Cell, 33, 85-93.
- Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol., 65, 499-560.
- Messing, J. and Vieira, J. (1982) Gene, 19, 269-276.
- Ollo, R. and Rougeon, F. (1983) Cell, 32, 515-523.
- Rechavi, G., Ram, D., Glazer, L., Zakut, R. and Givol, D. (1983) Proc. Natl. Acad. Sci. USA, 80, 855-859.
- Roth, M., Imanishi-Kari, T. and Rajewsky, K. (1979) Eur. J. Immunol., 9, 1004-1013.
- Rudikoff, S. and Potter, M. (1980) J. Immunol., 124, 2089-2092.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol., 143, 161-178.
- Schreier, P.H., Bothwell, A.L., Mueller-Hill, B. and Baltimore, D. (1981) Cell, 78, 4495-4499.
- Takemori, T., Tesch, H., Reth, M. and Rajewsky, K. (1982) Eur. J. Immunol., 12, 1040-1046.
- Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 1485-1489.
- White-Scharf, M.E. and Imanishi-Kari, T. (1981) Eur. J. Immunol., 11, 897-904.
- White-Scharf, M.E. and Imanishi-Kari, T. (1982) Eur. J. Immunol., 12, 935-942.