INTERACTION AND SEQUENCE DIVERSITY AMONG T15 $V_{\rm H}$ GENES IN CBA/J MICE

By STACY E. FERGUSON,* STUART RUDIKOFF,‡ AND BARBARA A. OSBORNE§

From the *Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, Massachusetts 01003; the [‡]Laboratory of Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; and the [§]Department of Veterinary and Animal Sciences and Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, Massachusetts 01003

The T15 V_{H} family in the inbred mouse encodes the majority of Ig heavy chain variable regions involved in the response to the haptenic determinant, phosphorylcholine $(PC)^1$ (1). Anti-PC antibodies also are elicited upon challenge with, and are protective against, the microbial pathogen Streptococcus pneumoniae (2). In most strains of inbred mice the T15 family consists of four genes that have been designated V1, V11, V13, and V3. The V1, V11, and V13 genes appear to be functional at the nucleotide level as they have the characteristic features of expressed genes (3, 4); while V3 is most likely a pseudogene (5). Sequence data (6-10) have demonstrated that it is primarily the V1 gene that is involved in the immune response to PC. An exception has been noted in a PC-binding CBA/J-derived hybridoma, 6G6 (11), in which the $V_{\rm H}$ sequence was found to share regions of identity with the V11, V13, and V1 genes of BALB/c (3). It was suggested that gene conversion among members of the T15 family was a likely explanation for the derivation of the 6G6 sequence. To unambiguously define the contributions of individual members of the family to the 6G6 sequence and to examine aspects of $V_{\rm H}$ evolution, we have characterized the germline T15 genes from CBA/J. In the present study we describe the isolation and sequence analysis of this family and suggest its likely involvement in multiple gene correction events occurring during the short evolutionary period representing the derivation of inbred strains.

Materials and Methods

 V_{μ} Probe. A cDNA clone encoding the T15 (V1) heavy chain (3) was generously provided by Dr. S. Crews, Division of Biology, California Institute of Technology, Pasadena, CA. For Southern blots and library screening a Pst I fragment containing relevant V_{u} sequence was isolated and labeled with [³²P]dCTP through random priming (12).

Genomic DNA Preparation and Southern Blots. CBA/J liver DNA was isolated as described by Maniatis et al. (13). Genomic Southern blots (14) were prepared and probed with the radiolabeled T15 V_H fragment. Hybridizations were performed at 65°C in 6X SSC buffer with washes in 0.1X SSC and 0.2% SDS at 65°C.

This work was supported by National Institutes of Health grant GM-36344 to B. A. Osborne. Address correspondence to B. A. Osborne, Dept. of Veterinary and Animal Sciences, University of Massachusetts, Amherst, MA 01003.

¹ Abbreviation used in this paper: PC, phosphorylcholine.

J. EXP. MED. © The Rockefeller University Press · 0022-1007/88/10/1339/11 \$2.00 1339 Volume 168 October 1988 1339-1349

Phage Genomic Libraries. A restricted CBA/J library was prepared by digesting liver DNA with Eco RI, followed by isolation of the 2–10-kb fragments on sucrose gradients and ligation into λ gtWES Eco RI arms. The 6.5-, 3.3-, and 3.0-kb fragments containing the V1, V3, and V11 genes, respectively, were isolated from a total of 12 positive clones from this library. Isolation of the 4.5-kb fragment containing the V13 gene was accomplished by digesting liver DNA with Eco RI, followed by agarose gel electrophoresis and electroelution of DNA corresponding to fragments 4.0-5.5 kb in size. These fragments were ligated into Eco RI arms of λ gt10 (Stratagene, La Jolla, CA) and packaged using commercial packaging extracts (Stratagene).

Library Screening. Phage were plated and plaque-lifts were taken using nylon filters (Hybond-N, Amersham Corp., Arlington Heights, IL). Filters were denatured, neutralized (15), and irradiated on a UV-transilluminator. Hybridizations were performed at 37°C in 50% formamide buffer (13) in the presence of radiolabeled T15 probe. Filters were washed in 0.1X SSC/0.2% SDS at 65°C and autoradiographed for 24 h. Positive clones were plaque purified and used for subsequent subcloning into M13.

M13 Subcloning. DNA from positive phage clones was isolated, digested with Eco RI, and electrophoresed through a 0.8% agarose gel for electroelution of the cloned inserts. Inserts were digested with Sau 3a and ligated into Bam HI-digested M13.

Dideoxy Sequencing. Sequencing of M13 phage clones containing Sau 3A inserts was performed using the dideoxy-method of Sanger et al. (16). Synthetic oligonucleotides were designed to prime at sequences immediately 5' to codon 1 and 3' to codon 101. These oligonucleotides were used to determine the DNA sequence of all four of the CBA/J T15 genes in both orientations. Sequencing of the 5' region of the CBA/J V1 gene was performed using the M13 universal primer.

Results

Southern Blot Analysis. Hybridization of genomic DNA from CBA/J mice with the T15 $V_{\rm H}$ probe under conditions of high stringency resulted in the identification of four major bands, as has been previously found in BALB/c and C57BL/10 (Fig. 1) (3, 4). Thus, the T15 family has not undergone expansion or contraction in the CBA/J strain. Phage clones containing the four major bands from CBA/J were isolated and characterized in the present study.

Sequence Analysis. Phage clones containing the four CBA/J T15 genes were subcloned into M13 vectors and sequences were determined as described in Materials and Methods. The aligned sequences of the four CBA/J coding regions as well as the previously reported 6G6 hybridoma are presented in Fig. 2. The 6G6 sequence differs from CBA/J V11 at one nucleotide position, establishing that 6G6 is encoded by this germline gene.

When the CBA/J T15 genes are compared with their BALB/c and C57BL/10 homologues (Fig. 3), three categories of substitution are found. First, there are single substitutions found uniquely in any one gene which are presumed to result from point mutations occurring after the separation of strains. 19 substitutions (circled in Fig. 3) are found that fall into this category and are exemplified by the A nucleotide in codon 23 of the BALB/c V13 gene. Secondly, there are a series of positions (marked by shading) in which a substitution is shared by two alleles but is not found in other family members. These substitutions most likely result from simple point mutations that occurred before the divergence of the two involved strains or alternatively could arise from parallel evolution. 15 positions are found corresponding to this category and are represented by the A nucleotide in codon 5 of the V13 gene in BALB/c and CBA/J. Of these 15 positions, 12 are shared substitutions between CBA/J and C57BL/10. This result is in agreement with known relationships of the origins of



FIGURE 1. T15 V_H genes in inbred mouse strains. High molecular weight DNA from BALB/c, C57BL/6, and CBA/J was digested with Eco RI, electrophoresed on an agarose gel, blotted, and hybridized with the T15 V_H probe. Molecular weights of the four homologous genes in each species are shown.

inbred strains (17) and presumably reflects point mutations in a common progenitor of CBA/J and C57BL/10. Two shared substitutions are found between CBA/J and BALB/c (codons 5 and 16) and one between C57BL/10 and BALB/c (codon 87). These substitutions are not explained by known strain derivation and may have arisen by parallel evolution which has been shown to occur frequently in C_K genes (18). Thirdly, there are differences that occur between allelic forms of a given gene in which the substituted nucleotide is found at the identical position in other T15 family members from the same strain. These (enclosed in boxes) presumably result from a gene correction event such as gene conversion and are enumerated in Table I. An example is codon 45 of the V1 gene where CTG is found in BALB/c and C57BL/10, while CCT occurs in CBA/J. The CBA/J V13 gene also has the sequence CCT at this position and thus is a potential donor of the CT substitution found in the CBA/J V1 gene. A similar pattern of substitution, in which appropriate donor sequences are found in other family members of the same strain, occurs at two positions in V1 and V13, four positions in V3, and five in V11.

There are a number of instances where putative gene corrections may be inter-

| c 6 | cac | CTC | | CTC | CTC | CAC | TCT | CCA | CCA | 10 | TIC | CTA | CAG | CCT | 666 | CCT | TOT | 676 |
|------------|-------|-----------|---------|-------|-------|---------|------|--------|----------|-------|-------|---------|--|--------|--------|--------|----------------|--------|
| 11 | ESS S | COLORS OF | 1010 | 010 | 100 | 000 | 1000 | COOM S | COLUMN 1 | and a | 100 | 1000 | Contra De la contr | 10.000 | 000 | 1001 | and the second | |
| i | 1000 | CONCELL. | - | | | | | | | | | | | 1 | | C | | |
| 13 | | | | | A | A | | | | | | | | | | -C- | | |
| 3 | | | | | | | | -A- | | | | | | | | | | |
| | | | | | | | | | | | | 20 | | | | | | |
| cc | ACT | CTC | TCC | TOT | | CCT | TCT | CCA | TTC | | TTC | ACT | CAT | TAC | TAC | ATC | ACC | TCC |
| 11 | AGT | CIC | 100 | 101 | UCA | L STORE | 101 | GUA | 110 | ACC | 110 | ACT | COAT I | THC. | IAC | Alto | AUC | 100 |
| ï | A | | | | | A | | G | | | £ | -G- | | · T · | | | GAG | |
| 13 | A | | | | | 1000 | 000 | 10.45 | 19993 | (cold | T | | | | | | | |
| 3 | A | | | | | A | | G | | | | | | -1- | | | -A- | |
| | | | | 40 | | | | | | | | | | 50 | | | | |
| 66 | CTC | ccc | CAC | CCT | cra | ccc | 440 | 600 | CTT | CAG | TCC | TTG | CCT | TIG | ATT | AGA | AAC | AAA |
| 11 | C. | 1000 | CAG | 003 | COA | 000 | AAG | GCA | and the | 0/10 | | ALC: N | 1000 | | | | | |
| ĩ | | | | G | | | | -G- | ·C- | | | A-1 | | GCA | -G- | | | |
| 13 | | | | -A- | | | | | -C- | | | | | 1000 | 1.44 | i Sand | 19.0 | |
| 3 | | T | | | | A | | | | | | | AG- | 1 | | | | |
| | | | | | | 60 | | | | | | | | | | 70 | | |
| 66 | GCT | AAT | GGT | TAC | ACA | ACA | GAG | TAC | AGT | GCA | TCT | GTG | AAG | GGT | CGG | TTC | ACC | ATC |
| 11 | | | | | | | | 102 | 1000 | 1000 | 100 | 1000 | 1000 | 1000 | Menter | 122/2 | 1000 | SEE: |
| 1 | | | -A- | 1 | | | | f | | | | | | | | | | 6 |
| 13 | 333 | 1042 | ê pix | 1.200 | 003 | - 203 | -035 | 1 | -C- | | ••• | 1 | | | | | | |
| 13 | | | | -1- | | | 0 | | | G | | A | | | | | | |
| | | | | | | | | 80 | | | | | | | | | | 90 |
| 66 | TCC | AGA | GAT | AAT | TCC | CAA | AGO | ATC | CTC | TAT | CTT | CAA | ATO | AAT | GCC | CTG | AGA | GCT |
| /11 | 300 | 6.500 | - bot | 1,000 | 1000 | 1000 | 1200 | 100 | 1000 | 1000 | 1,000 | | 2,000 | 1. 24 | | | | |
| 11 | | | (| -C- | | | | | | 0 | | 0 | | (| 1 | | | |
| /13 | | | • • • • | | | | -A- | •••• | | | ••• | • • • • | | (| 1A | | G | |
| 13 | | | | | | | | | | | | | | | 9.50 | C 7499 | -200 | - 1966 |
| | | | | | | | | | | 100 | | | | | | | | |
| 666 | GAG | GAG | AGT | GCC | ACT | TAT | TAC | TGT | GCA | AGA | GCT | r . | | | | | | |
| 111 | | | | | | | | | | | -A- | | | | | | | |
| 11 | | | - C- | | · ·T· | | | | | | -A- | | | | | | | |
| V13 | | •••• | | | | | | | -1- | -A- | -A- | | | | | | | |
| ٧3 | 22.4 | >: | 1.22 | 2 22 | 9.000 | - 222 | 22 | 2.03 | 1.40 | 200 | j. | 3 | | | | | | |

FIGURE 2. Nucleotide sequences of the CBA/J T15 genes. The nucleotide sequence of the four CBA/J germline genes are compared with that of the rearranged 6G6 gene. A hypothetical pathway for gene correction events involving the CBA/J V11 gene is shaded. Asterisks indicate an alternative pathway if correction occurred in the BALB/c V11 gene rather than the CBA/J V11 gene (see text).

preted to have occurred at either or both of two points in time (Table I). For example, at codon 24 the BALB/c V11 has an ACT while CBA/J and C57BL/10 have GCT. While the A nucleotide in the BALB/c codon may have originated from a correction event involving the BALB/c V1 gene, it is also possible that such a correction occurred between the V11 and the V13 genes in the CBA/J-C57BL/10 progenitor. Analogous situations are found at codons 6, 26, 32, 58, 86, and 87 and are marked by an asterisk in Table I. Alternatively, as in the case of the second category of substitutions described above, such sequences shared by CBA/J and C57BL/10 may result from point mutations in a common ancestor rather than by correction.

V3 Gene. The BALB/c V3 gene has been designated a pseudogene due to a number of structural defects, including a stop signal (TAG) at codon 47 and a 4-bp insertion at position 90 (5). These two anomalies presumably prevent the transcription of a full-length V region. Furthermore, the BALB/c V3 gene has a 7-bp deletion in the 23-bp spacer between the heptamer-nonamer splice recognition sites. Although it has been shown that the sequence of this 23-bp spacer region is not rigidly conserved, an absolute requirement for proper length has been demonstrated. Perlmutter et al. (4) have reported that the C57BL/10 V3 gene lacks the coding region defects found in BALB/c but retains the deletion in the 23-bp spacer, suggesting that this allele is also a pseudogene. We have found that the CBA/J V3 gene similarly has a 7-bp deletion in the 23-bp spacer, identical to that found in both BALB/c and C57BL/10. The CBA/J gene, like its C57BL/10 counterpart, lacks the coding region defects. Although there is no direct evidence to suggest that either the CBA/I or the C57BL/10 V3 alleles are pseudogenes, it is reasonable to presume they are nonfunctional due to the deletion in the 23-bp spacer.

| V 7/8/18 | | of of o | | - | 010 | | | | | 2 2 | | - | | | - | | - | - | 20 | | | - 0 - 0 | | | | | | 8 | - | | | 8 | |
|----------|-------|---------|--------|----------|-----|------|------|------|------|------|----------|--------|-----|------|------------|------|-----|-----|----------|------------|------|------------|------|-------|------|------------|------------|-----|---|-----|-----|-----|--|
| | | | | : : | | | ; | | - | - | | 5 | 1 | 3 | 3 , | 2 | 2 | | 5 | 2 | 5 | | - | 3 | - | - | Ĩ | Y | 3 | - | Y | 5 | |
| C5781 V | | | : | 1 | | | | : | | | | | - | | - | | | | | | | | | | | | | 1 | : | | | : | |
| BALB/C V | - | | ; | : | 1 | - | 1 | : | : | : | | | 2 | | 8 | | | | | | | | | | | | | | | | | : | |
| CBA/J V | = | | : | : | : | | : | ; | ; | : | : | : | 1 | | | | 1 | 8 | | | | | | | _ | | | ļ | | | | | |
| C5781 V | - | | : | : | : | | : | ; | : | ; | : | : | : | : | 1 | : | ; | | : | 1 | : | : | ; | | - | ł | 1 | , i | : | | | : | |
| BALB/C V | . 51 | : | : | : | - | Ģ | : | ; | : | ; | : | : | : | : | цų. | - | : | : | : | : | : | | : | 1 | : | 1 | 1 | ÷ | : | + | : | : | |
| CBA/J V | - 51 | | : | : | i. | + | 1 | ; | : | ; | : | : | : | : | ný. | - | : | : | : | 1 | : | : | ; | : | : | 1 | 105 | ÷ | : | + | ; | : | |
| C5781 V | - 51 | | : | : | | Ż | ; | ; | : | : | : | : | : | • | 8 | 1 | : | : | : | : | : | : | ; | : | : | 1 | 1 | ÷ | ; | | : | : | |
| BALB/C V | | : | : | : | : | 0 | ; | ; | : | : | : | : | : | : | - | - | : | : | : | - | : | ·e | : | - | : | 1 | | , i | : | | - | : | |
| C8A/J V | | : | : | : | : | ÷ | 1 | , | : | : | : | : | : | : | - | - | : | : | ; | i | : | : | - | - | ł | 1 | - | , i | : | - | : | : | |
| C5781 \ | . 5 | : | : | : | : | 9 | : | Like | : | : | : | : | 1 | 1 | - | 1 | ; | : | : | : | : | : | : | : | : | 1 | ł | ÷ | : | -; | : | : | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | 2 | | | | | | | | | |
| | | 100 100 | -10 | | | | | | 0 | | E. | | | | | | | | | | | | 1 | • | | | | | | | | | |
| CRAVI | | | ; ; | ; ; | 3 | 1 | 5 | | | 1 | | | | 3 | 3 | | | | | | | | | | Y CA | DI I | No. | GCA | 5 | 5 | | 5 | |
| C5781 V | | | : | : | 0 | 100 | : | ; | : | | | | | | | | | | | | | | | | | | | | | | | | |
| BALB/C V | A 11 | | : | ; |): | | : | e | : | : | 1 - | : | | 4 | the second | -1-1 | : | ; | : | | | | | | | | | | | | | 8 | |
| CBA/J | A 111 | | : | : | : | ; | : |): | : | | | : | | | | | : | 1 | | | | | | | | | | | | | | | |
| CS781 | | | : | : | : | G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | Ð | | | 0 | | | | - • | , c | - | - | : | : | : | | | | : | : | : | | ! | : | : | : ' | : | : | |
| 1/01/0 | | | | | | (| | | 2 | | 5 | | ÷ . | : | Ē | - | : | : | ; | | | | - | : | : | | 1 | : | ÷ | Ŧ | - | : | |
| C/YS | 2 | | : | : | : | 3 | : | | : | | 5 | - | ÷ | : | E - | - | : | : | : | : | : | | : | : | : | - | ķ 1 | ÷ | : | : | : | : | |
| C5781 | × 113 | : | : | : (| : | : | : | | : | | 5 | : | ÷ | : | 11 - | | : | : | : | ÷ | : | | : 0 | : | : | - | ¢1 | : | : | : | : | : | |
| SALB/C | × | | 1 | 3 | : | : | ľ | 1. | : | | - | 0 | 0 | : . | - 11 | | : | ÷ | : | i | : | | 0 | : | : | • | 1 | : | : | : | : | : | |
| CBA/J | × | | : | in | : | : | : | i | : | | - | : | - | Ś | | | : | : | : | : | ; | | : | : | 1 | : | ł | e | : | : | : | : | |
| C5781 | • | | : | ie | : | : | : | į. | : | : | - | : | ÷ . | 9. 9 | | | : | : | : | : | : | | : . | : | - | - | 1 | : | : | :. | : | : | |
| | | | | | | | 5 | | | | | | | | | 1 | | | | | | | | | - | a - | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | 02 | 227 | | | | | | | | | 9 | | | | | | | | | • | • | | | | | | | | | 001 | | |
| | | | | - | | - | - | : | | | | 1 | | - | | | 5 | | | | • • | | | | - | | | | : | | | | |
| | | | | : | 1 | | | | | | | | | | 5 | | 1 | ; ; | | 5 | 2 | | | | 3 | | | | 2 | 5 | 5 | : : | |
| 19731 | | | | | | | | | | | | | | | | | - | : | - | | | | | | | | | | | | | | |
| | | | - | | | | | | | | | | | | | | | F | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | 4 | - | | | | | | | | | | | | | | | |
| | | | | | | | • | | | | | | | | | | | 4 | | | 8 | | | | | | | | | | | | |
| | | | • | | | | | | 0 | | | | | | | | 1 | 5 | | | | | | | | | | | | | | | |
| | 2 2 | | • | | | | | |) | | | | | | | | 1 | | | 2 | | | | | | | | | | | | | |
| | | | • • | | | | | | | | | | | | | | • • | | | | | | | • | | • • | | | | | | (| |
| | 2 . | | ÷ | | : | : | - | | | | | | | : | 1 | | - 6 | | : | | | 1 | : | | : | | • | : | : | ÷- | - | 5) | |
| BALB/C | | | ÷ | - | : | : | 7 | | : | : | : | : | : | : | - | | | - | : | ė | -ACA | | : | | : | ÷ | 1 | : | : | : | : | : | |
| CBA/J | | : | ÷ | : | : | : | 7 | | ; | : | : | : | : | - | - | - | ÷ | - | : | : | : . | • | : | | : | ÷ | : | ÷ | ÷ | : | : | : | |
| C5781 | | : | ÷ | : | : | : | 7 | ł | : | ÷ | : | ÷ | : | : | 1 | ÷ | ÷ l | - | : | | 0 | : | : | | : | ÷ | ÷ | : | : | : | : | : | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Ģ | | ¢ | (| | | | | Ê | ĥ | | Ê | 1.4 | | 1 | 5 20 | 10 | ç | и 1 | | 1 | Ē | | | 1 | | | ŝ | | | | | |
| | Ξ. | GUR | n N | <u>.</u> | E. | par | 1201 | 101 | Ω. | | ູ່. ເ | r n | ŝ | an. | י פ | | | ⊇`' | <u>.</u> | gen Sen | Si ' | р. Ц | uns. | Deg - | 2 | Ξ. | | ġ | | | | | |
| | Ĕ | ons a | ure c | irc | g | dit | fere | nce | S SI | nare | å b | ž L | 0 M | alle | s | but | bou | ē | pur | E | oth | er F | ami | ž | ner | npe npe | S | гe | | | | | |
| | 4° | adeo | an | Ę | ote | 'nti | ale | ene | 5 | rre | "tio | ns : | are | ĝ | red. | | | | | | | | | | | | | | | | | | |
| | 1 | aucr | 1, G. | 2 | ś. | 1111 | 5 | i) | 3 | 2 | 5 | | | ŝ | į | | | | | | | | | | | | | | | | | | |

FERGUSON ET AL.

| V1 | | Codon 45 | Codon 86 | V13 | Codon 6* | Codon 58* |
|------|----------------|----------------|------------------|------------|-------------|---------------------|
| DATD | V1 | OTO | 4 A T | DAT D 3719 | | T + T |
| DALD | VI V11 | | AAI | BALB VIS | GAG | IAI |
| | V 1 1 V 1 2 | 1 CT | | V 3 | | C |
| | V15 V2 | -UI T | | VI VI | A | |
| | V 3 | 1 | | VII | | C |
| CBA | V1 | -CT | - ~ Č | CBA V13 | Å | Č |
| | V11 | T | | V 3 | A | - TC |
| | V13 | - CT | C | V1 | A | |
| | V3 | T | | V11 | | C |
| C57 | VI | | | C57 V13 | A | C |
| | V11 | T | | V3 | | C |
| | V13 | $-\mathbf{CT}$ | C | V1 | A | |
| | V3 | T | C | V11 | | C |
| V11 | | Codon 24* | Codon 26* | Codon 49 | Codon 50 | Codon 86-87* |
| | 1711 | ↓ ↓ CTT | | | | + + |
| BALB | V11 | ACT | GGG | GGT | TIT | AAC ACC |
| | V13 | G | A | -0- | G | |
| | V3 | 1 | | -0- | G | A |
| | VI | | | - C - | GCA | T G |
| CBA | V11 | * G | Å | -Č- | G | T G |
| | V13 | G | A | - C - | G | |
| | V3 | | | A | | |
| | V1 | | | - C - | GCA | G |
| C57 | V11 | G | A | | | T G |
| | V13 | G | A | - C - | G | |
| | V3 | | | | | A |
| | V1 | | | - C - | GCA | T G |
| V3 | | Codon 6* | Codon 32* | Codon 50 | Codon 86-87 | |
| | | ~ · • | t | + | | |
| BALB | V3 | GAG | TAC | TTG | AAC ACA | |
| | VI | A | - 1 - | GCA | 1 G-C | |
| | VII | | | 1 | C | |
| | V13 | | | | | |
| СВА | V3 | Å | - Ť - | T | TG-C | |
| | V1 | A | - T - | GCA | G-C | |
| | V11 | | | G | T G-C | |
| | V13 | A | | G | C | |
| C57 | V٩ | | - T - | T | | |
| 357 | Vi | A | - <u>-</u> - T - | GCA | T G - C | |
| | VII | / 1 | | T | T G-C | |
| | V13 | A | | | C | |

TABLE I Potential Sites of Gene Corrections

Arrows indicate the sites of gene correction. * Indicates a position where gene correction may have occurred in either BALB/c or the CBA/J-C57BL pro-genitor. Alternatively, in the case of CBA/J-C57BL, the substitution pattern could be explained by a point mutation in the common progenitor rather than a correction event.

FERGUSON ET AL.

| | LEADER |
|--------|--|
| CBA/J | TTCACATTCAGAAATCAGCACTCAGTCCT GTCACTATCAAGTTGTGGTT |
| BALB/c | TTT |
| C57/BL | TA |
| CBA/J | AAACTGGGTTTTTCTTTTAACACTTTTACATGGTAATTTATGGGCAAAAA |
| BALB/c | |
| C57/BL | |
| CBA/J | GAGTTCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT |
| BALB/c | ТТ. |
| C57/BL | TT |
| CBA/J | GTGTGTGTGTGTGTGTGTGT |
| BALB/c | GCATATGTG GCATATGTG |
| C57/BL | GTGTGTGTGTGTGTGTGTGTGTGTGCATATGTG |
| CBA/J | AACAGGGATAGAAGATTAGTAGATGTGTATG ACCAGCTTTCTCTAC |
| BALB/c | TGTTTT |
| C57/BL | TGT |
| 031700 | LEADER One |
| CBA/J | TATTGCAGGTATCCAGTGT |
| BALB/c | |
| C57/BL | |

FIGURE 4. Comparison of 5' sequences of V1 genes. Boxed regions encode the leader sequence exons and the hatched region denotes the first codon of the V1 coding region. Shaded areas indicate GT-dinucleotide repeats.

V1 5' Sequences. The sequence just upstream of the CBA/J V1 gene including untranslated region as well as the leader sequence has been determined and compared with the 5' regions of the C57BL/10 and BALB/c V1 genes (Fig. 4). The intron within the leader sequence contains a long GT-dinucleotide repeat that is found in all three strains, although the length of this repeat is variable. The potential relevance of this repeat is addressed in the Discussion.

Discussion

Previous studies by Clarke et al. (11, 19) have described a CBA/J V_{H} sequence (6G6) that appeared to be hybrid in nature, containing sequence elements found in several genes of the T15 family. It was proposed that the 6G6 sequence arose by gene conversion involving different members of this V_H family. Gene conversion, the process by which one nucleotide sequence is "corrected" by a homologous, but nonidentical, sequence by excision and repair without the exchange of genetic material, has been precisely defined at both the mitotic and meiotic levels in fungi and yeast where individual products of meiosis may be examined (20-23). Although a similar analysis of meiotic products is not possible in higher eukaryotic cells, this process has been implicated in the evolution of mammalian multigene families such as the globins (24), Ig γ chains (25), human V_K light chains (26, 27), mouse V_H segments (28), TCR constant regions (29), and genes of the MHC (30). While it is likely that this process is similar in fungi and mammals, the inability to analyze the phenomenon at the same level in more highly evolved cells remains. Therefore we have chosen to refer to our observations in the T15 V_H gene family as gene correction rather than use the term gene conversion precisely defined only in yeast and fungi.

In the case of 6G6, the V_{H} sequence suggested to arise by gene conversion was derived from a CBA/J hybridoma and thus was the product of a rearranged gene. Therefore it could not be determined whether the putative conversion occurred during somatic diversification or in the CBA/J germline. To both understand the mechanisms contributing to the evolution of multigene families and, more specifically, to document the 6G6 gene conversion, sequences of the CBA/J T15 V_H germline genes were determined.

From Fig. 2, it can be seen that the 6G6 sequence is, with the exception of a single nucleotide, exactly encoded by the CBA/J V11 gene. This gene has regions of nucleotide sequence that are more homologous to CBA/J V13 and V3 genes than to corresponding regions in the BALB/c V11 gene. The CBA/J V11 gene is, in fact, a variation of the BALB/c allele that may be generated by incorporating sequences from other members of the CBA/J T15 $V_{\rm H}$ family. Thus, these results support the proposition of Clarke et al. (11, 19) that 6G6 arose by gene correction and further demonstrate that the correction occurred in the CBA/J germline. Other examples of putative gene corrections were found to occur in the CBA/J V1, V3 and V13 germline genes (Table I), which suggests that gene correction may be a frequent event in the evolution of this multigene family as, in these instances, corrections have occurred subsequent to the separation of BALB/c and CBA/J. These results do not, in any way, preclude the possibility that, in other instances, gene correction may occur at the somatic level.

An alternative explanation for the derivation of such sequences would be homologous recombination. In the case of the CBA/J V11 gene, five crossovers within 300 bp would have had to occur to give rise to the observed V11 sequence. Furthermore, unequal crossing over is unlikely to be involved in the diversification of the T15 family as the number of family members remains constant in all three strains (Fig. 1). Therefore, we favor the interpretation that these sequences have been generated by a gene correction mechanism such as gene conversion. The identification of appropriate donor sequences in other family members provides a source of DNA necessary for the origin of the proposed converted sequences.

In addition to gene corrections, two other categories of substitutions were observed in the analysis of CBA/J T15 V_{H} families (Fig. 3). The first category was simple single base substitutions presumed to be the result of point mutation. It is interesting to note that 9 of these 19 substitutions are found in the CBA/J V3 gene which, similar to its counterparts in BALB/c and C57BL/10, appears to be a pseudogene based on a 7-bp deletion in the 23-bp spacer between the heptamer/nonamer sequences immediately 3' to the V gene. It might be expected that a pseudogene would contain more point mutations than other potentially functional members of a multigene family, as there would be no selective pressures acting on this sequence. The fact that the majority of such mutations are observed in the pseudogene supports such a conclusion. The second category of substitutions involved a unique nucleotide shared by two alleles but not found in any other family member. Such substitutions are most common to C57BL/10 and CBA/J, as might be expected from their proposed origins (17). The simplest interpretation of events giving rise to this category is that a common progenitor of C57BL/10 and CBA/J either acquired a mutation or underwent a gene correction event giving rise to the shared nucleotide found at the particular position in these two strains. In the three cases where CBA/J and BALB/c or C57BL/10 and BALB/c share nucleotides, parallel evolution would appear to be the most likely explanation.

As part of the analysis of the CBA/J T15 family, sequences immediately 5' of the V1 gene also were determined. Previously published data for this region in BALB/c and C57BL/10 have revealed a long GT-dinucleotide repeat constituting a large portion of the leader peptide intron. This repeat has also been detected upstream of

the CBA/J V1 gene (Fig. 4). Although the short sequence CATA appears to interrupt the GT repeat in the BALB/c and C57BL/10 sequences, these nucleotides do not change the alternating purine and pyrimidine nature of the region. Differences in the lengths and minor sequence variation of this repeat between strains could be due to slippage mispairing during replication (31). The GT repeat is found frequently in eukaryotic genomes (32). While its exact function(s) are not clear, some lines of evidence suggest that it may have enhancer activity (33). Furthermore, purine/pyrimidine dinucleotide repeats are believed to have left-handed, Z-DNA forming potential (34). It is tempting to speculate that a sequence with this alternative conformational capability may exert some effect on the expression, rearrangement, or evolution of the T15 V1 gene, particularly since, in this case, the unusual sequence lies quite close to the V1 coding region. These possibilities are currently under investigation.

The above studies demonstrate that gene correction may contribute, in conjunction with the other described mechanisms, to the generation of diversity in the antibody repertoire. The potential biological effects of such a process are several. First, new alleles may be generated by the inclusion of sequence segments from different family members. These sequences can be donated either by functional or pseudogenes providing, in the latter case, a means for rescuing useful sequence information from genes which have incurred crippling mutations or, alternatively, for correcting deleterious mutations by replacement with nondefective sequences. Secondly, intrafamily homology can be maintained by the effective homogenization resulting from such a process. Thirdly, substitutions can be spread to other family members without the necessity of the same rare event occurring multiple times. In the future it will be important to determine if, and how frequently, gene correction may occur somatically to further amplify diversity within an individual. It is reasonable to expect that such a process might contribute further Ig diversity during somatic amplification of a particular B cell clone since chicken Ig λ light chains have been shown to use a single functional variable region gene that undergoes multiple correction events with many pseudogenes to diversify its light chain repertoire (35, 36).

Summary

Nucleotide sequences of the four genes composing the T15 heavy chain variable region (V_{H}) family of the CBA/J mouse have been determined. Comparison of these sequences with their published BALB/c and C57BL/10 homologues reveals that nucleotide differences found between given alleles of two strains, i.e., CBA/J and BALB/c, are observed in other family members of the same strain. We suggest that these patterns of sequence variation are most readily explained by gene interaction (conversion). Additionally, the sequence of a CBA/J hybridoma, 6G6, proposed to have been generated by gene conversion, is directly encoded by the CBA/J V11 gene indicating that the putative conversion has occurred meiotically in the germline. These results are consistent with the premise that gene correction is occurring frequently among members of this family and that such processes may contribute significantly to the evolution of Ig variable region genes even in the relatively short time frame of inbred strain derivation.

We thank Dr. Stephen Clarke for his insight, help, and encouragement during this project; Dr. Richard Goldsby for useful criticism of the manuscript; and Ms. Lisa Scott for expert editorial assistance.

Received for publication 8 June 1988.

References

- 1. Claffin, J. L., and S. Rudikoff. 1977. Uniformity in a clonal repertoire: a case for a germline basis of antibody diversity. *Cold Spring Harbor Symp. Quant. Biol.* 41:725.
- Briles, D. E., C. Forman, S. Hudak, and J. L. Claffin. 1982. Antiphosphorylcholine antibodies of the T15 idiotype are optimally protective against Streptococcus pneumoniae. J. Exp. Med. 156:1117.
- Crews, S., J. Griffin, H. Huang, K. Calame, and L. Hood. 1981. A single V_R gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with class of antibody. *Cell.* 25:59.
- Perlmutter, R. M., B. Berson, J. Griffin, and L. Hood. 1985. Diversity in the germline antibody repertoire: molecular evolution of the T15 V_H gene family. J. Exp. Med. 162: 1998.
- Huang, H., S. Crews, and L. Hood. 1981. An immunoglobulin V_H pseudogene. J. Mol. Appl. Genet. 1:93.
- 6. Claffin, J. L., and S. Rudikoff. 1979. Structural evidence for a polymorphic or allelic form of the heavy chain variable region. J. Immunol. 122:1402.
- Siu, G., E. A. Springer, H. Huang, L. E. Hood, and S. T. Crews. 1987. Structure of the T15 V_H gene superfamily: identification of immunoglobulin gene promoter homologies. J. Immunol. 12:4446.
- 8. Gearhart, P., N. D. Johnson, R. Douglas, and L. E. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.).* 219:29.
- Perlmutter, R. M., S. T. Crews, R. Douglas, G. Sorenson, N. Johnson, N. Nivera, P. J. Gearhart, and L. Hood. 1984. The generation of diversity in phosphorylcholine-binding antibodies. *Adv. Immunol.* 35:1.
- Clarke, S. H., J. L. Claffin, M. Potter, and S. Rudikoff. 1983. Polymorphisms in antiphosphocholine antibodies reflecting evolution of immunoglobulin families. J. Exp. Med. 157:98.
- 11. Clarke, S. H., J. L. Claffin, and S. Rudikoff. 1982. Polymorphisms in immunoglobulin heavy chains suggesting gene conversion. *Proc. Natl. Acad. Sci. USA*. 79:3280.
- 12. Feinberg, A. P., and B. Vogelstein. 1983. A technique for labeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6.
- 13. Maniatis, T., E. F. Fritsch, and J. Sandbrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
- 14. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503.
- 15. Benton, W. D., and R. W. Davis. 1977. Screening recombinant clones by hybridization to single plaques in situ. Science (Wash. DC). 196:180.
- 16. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143:161.
- 17. Morse, H. C. 1978. Origins of Inbred Mice. Academic Press, New York. 719 pp.
- Jovin-Marche, E., A. Cuddihy, S. Butler, J. N. Hansen, W. N. Fitch, and S. Rudikoff. 1988. Modern evolution of a single copy gene: The C_κ locus in wild mice. *Mol. Biol. Evol.* In press.
- 19. Clarke, S. H., and S. Rudikoff. 1984. Evidence for gene conversion among immunoglobulin heavy chain variable region genes. J. Exp. Med. 159:773.

FERGUSON ET AL.

- 20. Fogel, S., R. Mortimer, K. Lusnak, and F. Tavares. 1978. Meiotic gene conversion: a signal of the basic recombination event in yeast. Cold Spring Harbor Symp. Quant. Biol. 43:1325.
- 21. Klein, H. L., and T. D. Petes. 1981. Intrachromosomal gene conversion in yeast. Nature (Lond.). 289:144.
- Jackson, J. A., and G. R. Fink. 1981. Gene conversion between duplicated genetic elements in yeast. Nature (Lond.). 292:306.
- 23. Nagylaki, T., and T. D. Petes. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics*. 100:315.
- Slighton, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal γ and α globin genes: complete nucleotide sequences suggest that DNA can be exchanged between duplicated genes. *Cell.* 21:627.
- 25. Ollo, R., and F. Rougeon. 1983. Gene conversion and polymorphism: Gene conversion and generation of mouse immunoglobulin gamma 2a chain alleles by differential gene conversion by gamma 2b chain gene. *Cell*. 32:515.
- 26. Bentley, D. L., and T. H. Rabbitts. 1981. Evolution of immunoglobulin V genes: Evidence indicating that recently duplicated human V_{κ} sequences have diverged by gene conversion. *Cell.* 32:181.
- Akimenko, M. A., B. Mariame, and F. Rougeon. 1986. Evolution of the immunoglobulin kappa light chain locus in the rabbit: evidence for differential gene conversion events. *Proc. Natl. Acad. Sci. USA.* 83:5180.
- Krawinkle, U., G. Zoebelein, M. Bruggemann, A. Radbruch, and K. Rajewsky. 1983. Recombination between antibody heavy chain variable-region genes: evidence for gene conversion. *Proc. Natl. Acad. Sci. USA*. 80:4997.
- 29. Jovin-Marche, E., M. Heller, and S. Rudikoff. 1986. Gene correction in the evolution of the T cell receptor β chain. J. Exp. Med. 164:2083.
- 30. Geliebter, J., and S. Nathanson. 1987. Recombination and the concerted evolution of the murine MHC. *Tiends Genet.* 3:107.
- 31. Streisinger, G., Y. Okada, J. Emrich, J. Newton, J. Tsugita, A. Terzaghi, and M. Inouye. 1966. Cold Spring Harbor Symp. Quant. Biol. 31:77.
- Morris, J., S. R. Kushner, and R. Ivarie. 1986. A simple repeat, poly(dT-dG):poly(dCdA), common to eukaryotes is absent from eubacteria and archaebacteria and rare in protozoans. *Mol. Biol. Evol.* 3:343.
- 33. Hamada, H., M. Seidman, B. H. Howard, and C. M. Gorman. 1984. Enhanced gene expression by the poly(dT-dG):poly(dC-dA) sequence. *Mol. Cell. Biol.* 4:2622.
- Nordheim, A., and A. Rich. 1983. The sequence (dC-cA)n:(dG-dT)n forms left handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA. 80:1821.
- 35. Reynaud, C.-A., V. Anquez, H. Grimal, and J.-C. Weil. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell.* 48:379.
- 36. Thompson, C. B., and P. E. Neiman. 1987. Somatic diversification of the chicken light chain gene is limited to the rearranged variable gene segment. *Cell.* 48:369.