ANALYSIS OF A HUMAN V_β GENE SUBFAMILY

By GERALD SIU, ERICH C. STRAUSS, ERIC LAI, and LEROY E. HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

The T cell antigen receptor consists of two chains, denoted α and β, each of which is composed of a variable (V) region and a constant (C) region (reviewed in 1 and 2). The β chain consists of a V region of ~120 amino acids and a C region of ~170 amino acids (3, 4). The V_β region is encoded by three gene segments, V_β, D_β, and J_β, that are joined together during T cell development to generate a complete V_β gene (5, 6). DNA rearrangement recognition signals that mediate variable gene formation are located 3′ to the V_β gene segments, 5′ to the J_β gene segments, and on both sides of the D_β gene segments; these are similar to the rearrangement recognition signals of immunoglobulins, and consist of conserved heptamer and nonamer sequences that are separated by variable spacer sequences of either 12 bp or 23 bp (5–11).

The germline V_β gene segments in the mouse genome are grouped into many single- or low-copy subfamilies of closely-related gene segments, each member ≥75% homologous, at the DNA level, to the other members (12–14). Of the 14 mouse V_β gene segment subfamilies analyzed to date, 12 contain only one member, and two contain three members (12–14). Human V_β gene segments appear to be organized in a similar fashion. The 15 V_β subfamilies consist of one subfamily of six members, one with five members, one with four members, two with three members, four with two members, and six with one member (15). In contrast, the immunoglobulin and α chain V gene segments can be grouped into subfamilies that range in size from 4 to over 50 members (16–20).

In the course of our study of the human V_β gene segments, we identified one subfamily, originally denoted $V_{\beta M3}$ and now referred to as $V_\beta 8$, that contained five members and is one of the largest V_β subfamilies reported to date (6). This subfamily provided a unique opportunity to study the structure, organization, and evolution of V_β subfamilies. We have isolated cosmids and λ clones containing the five members of this subfamily, and each was characterized by subcloning and sequence analysis. In addition, we have used the field-inversion gel electrophoresis technique (21 and E. Lai and L. E. Hood, manuscript in preparation) to analyze the genomic organization and linkage of the $V_\beta 8$ subfamily.

Materials and Methods

Genomic Blots. Genomic blots were carried out as previously described (22, 23). The YT35 V_β region-specific DNA probe for the hybridizations was labeled with ^{32}P using the protocol of Rigby et al. (24). Hybridizations were carried out at 68°C for 24 h as described (22, 23). After hybridization, the filters were washed three times with 3× SSC

This work was supported by grants from the National Institutes of Health, Bethesda, MD, and from T Cell Sciences, Inc. E. Lai is a Leukemia Society of America fellow.

for 20 min each followed by three washes with 1× SSC for 20 min each, at 68°C. The filters were then exposed to Kodak XAR-5 film overnight at −70°C with an intensifying screen.

Construction and Screening of Genomic Libraries. A human sperm library in the pTL5 cosmid vector was provided by Lance Fors (California Institute of Technology, unpublished data), and screened with the YT35 V_β region–specific probe as previously described (23). Partial λ libraries were constructed from human sperm DNA into the vector λgt7lac5 and screened with the YT35 V_β probe as described (25).

Restriction Endonuclease Mapping. Mapping of different restriction endonuclease sites was accomplished by comparison of single and double digests of restriction endonucleases of the genomic clones. In addition, a novel restriction endonuclease mapping technique (Sun, Y., and L. Hood, manuscript in preparation) based on the technique of Smith and Birnstiel (26) was used. This technique depends on the presence of a Sal I, Cla I, or Nru I site present in the cloning vector and not in the insert DNA. In our case, a cosmid was digested to completion with Sal I, and aliquots were digested with different concentrations of another enzyme that would permit only incomplete digestion. A range of the partial-digestion reactions were subjected to electrophoresis on agarose horizontal slab gels. After electrophoresis, the gels were denatured, neutralized, blotted with nitrocellulose, and the nitrocellulose was baked as described for genomic blots (23). Two synthetic oligonucleotide hybridization probes, corresponding to the regions of DNA on either side of the Sal I site in the cosmid vector, were used for screening. The oligonucleotides were labeled with ^{32}P using polynucleotide kinase according to the protocol of Maxam and Gilbert (27), and used as hybridization probes. The probes will hybridize to each partial-digestion product that contains the vector sequence adjacent to the Sal I site on one end and a recognition site for the restriction enzyme tested on the other end. The hybridization will thus result in a ladder of bands, each corresponding to a different restriction enzyme site in the clone. In this manner, the restriction sites for each restriction enzyme were mapped with respect to the Sal I site of the vector. Both oligonucleotides were necessary to map accurately all of the restriction sites in the clones.

Subcloning and DNA Sequencing. Subcloning was carried out using the procedures described previously (23). Sequencing was carried out using the dideoxynucleotide sequencing technique as described by Strauss et al. (28).

Field-inversion Gel Electrophoresis. Genomic DNAs from human sperm and the HeLa cell line were digested with restriction enzymes and subjected to field-inversion gel electrophoresis as described (21 and E. Lai and L. E. Hood, manuscript in preparation). The gels were run with a switch-interval ramp in which the forward migration interval varied linearly from 4 to 15 s. The applied voltage gradient was 10 V/cm, and the running time was 20 h. After electrophoresis, the gels were denatured and neutralized, and the DNA was transferred to nitrocellulose as described for genomic blots (22, 23).

Results and Discussion

Isolation of Genomic V_β Clones. A V gene segment–specific probe was isolated from the YT35 β chain cDNA clone (3). With this probe, five hybridizing bands were detected on a genomic blot of human DNA. These five bands corresponded to five different V_β gene segments, which were denoted $V_\beta 8.1$, $V_\beta 8.2$, $V_\beta 8.3$, $V_\beta 8.4$, and $V_\beta 8.5$ (Fig. 1). These correspond to the $V_{\beta M3\text{-}1}$, $V_{\beta M3\text{-}2}$, $V_{\beta M3\text{-}3}$, $V_{\beta M3\text{-}4}$, and $V_{\beta M3\text{-}0}$ gene segments described earlier (6). A human cosmid library constructed from sperm DNA was screened with the YT35 V_β hybridization probe. Two cosmids containing three of the members of the $V_\beta 8$ subfamily were subsequently isolated: the cosmid clone H7.1 contains the $V_\beta 8.1$ and $V_\beta 8.2$ gene segments, and the cosmid clones H9.1 and H18.1 contain the $V_\beta 8.4$ gene segment. To characterize the other members of the $V_\beta 8$ subfamily, λgt7 libraries were constructed from 4.4 and 8.5 kb fragments of Eco RI–digested genomic

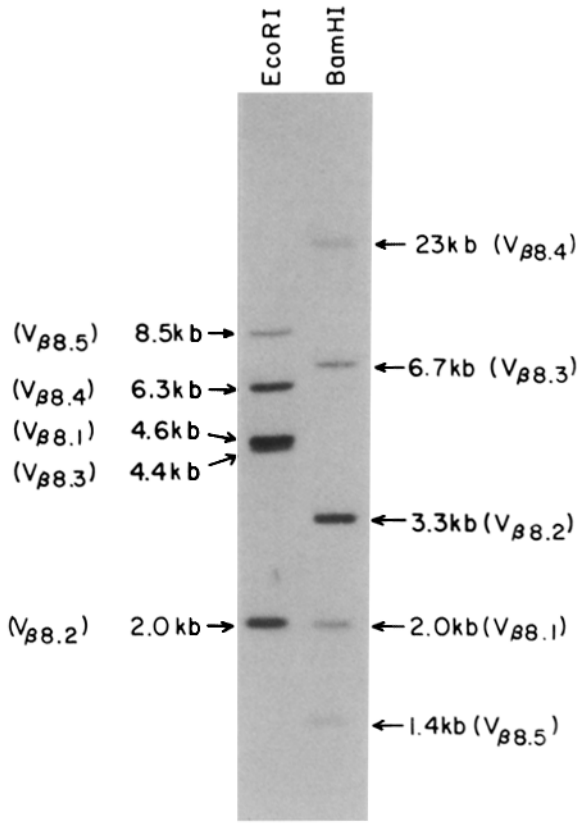

FIGURE 1. Genomic blots on human germline DNA with the YT35 V region–specific probe. The DNA samples were digested with the restriction enzymes Eco RI or Bam HI. The sizes of the bands are indicated, along with the corresponding V_β gene segment.

DNA constituting the size range of DNA fragments that contain the $V_\beta 8.3$ and $V_\beta 8.5$ gene segments, respectively. As with the cosmid library, these λ libraries were screened with the YT35 V_β-specific probe, and λ clones containing the remaining members of the $V_\beta 8$ subfamily were isolated: the λ clone λVB8.3 contains the $V_\beta 8.3$ gene segment, and the λ clone λVB8.5 contains the $V_\beta 8.5$ gene segment. The five members of the $V_\beta 8$ subfamily were analyzed by restriction mapping, subcloned, and the coding and flanking regions were sequenced (Figs. 2 and 3).

The $V_\beta 8$ Subfamily. The $V_\beta 8$ subfamily contains five members and accordingly is one of the largest V_β gene subfamilies characterized to date (12–15). Each gene segment consists of two exons: the first is 49 bp and encodes most of the leader peptide, and the second is 295 bp and encodes the remaining five amino acids of the leader and the coding region of the V_β gene segment. The two exons are separated by an intron of ~100 bp. The 3′ flanking regions of these gene segments contain the conserved heptamer and nonamer sequences and the 23 bp spacer sequence that are characteristic of one of the DNA rearrangement recognition signals for V_β gene formation (5–11).

The $V_\beta 8.1$ gene segment is identical to the V_β segments of the YT35, REX,

TABLE I
Homology Matrix of V_β Gene Segments

	V_β8.1	V_β8.2	V_β8.3	V_β8.4	V_β8.5	Murine V_β11
V_β8.1	—	97.9	77.7	70.2	66.0	69.9
V_β8.2	98.2	—	75.5	69.2	64.9	69.9
V_β8.3	82.1	82.1	—	63.8	60.6	66.7
V_β8.4	79.2	79.6	77.1	—	59.6	62.4
V_β8.5	76.4	75.4	72.9	76.1	—	59.1
Murine V_β11	77.4	77.8	77.4	74.9	71.3	—

* Numbers above the diagonal designate the percentage homology of sequences on the *x* and *y* axes when compared at the protein level; numbers below the diagonal show percentage homology at the DNA level.

and JM β genes (3, 29, 32), and accordingly represents the germline V_β gene segment that encodes these V_β genes. These data indicate that somatic hypermutation has not occurred in these tumors, and are consistent with other studies in which there has been no demonstration of somatic hypermutation in the β chain of the T cell receptor (5, 30, 31). The YT35 V_β gene is 98% identical in DNA sequence to the V_β8.2 gene segment, 82% identical to V_β8.3, 79% identical to V_β8.4, and 76% identical to V_β8.5 (Table I). The V_β8.1, V_β8.2, and V_β8.3 gene segments appear to be functional. In contrast, the V_β8.4 and V_β8.5 gene segments have in-frame stop codons; thus, two of the five members of this family appear to be pseudogenes. Two of the gene segments, V_β8.1 and V_β8.2, are separated by 3 kb of DNA and are in the same transcriptional orientation (Fig. 2*A*). These V_β gene segments share extensive sequence similarity both in their coding and flanking regions (Fig. 3). Presumably, one of these gene segments arose from the other by a recent gene duplication event.

A clone-specific mAb called Ti_{3A}, reactive to an epitope expressed by the V_β8 subfamily, was used to identify T cell clones that used these gene segments (32). One of the T cell clones apparently rearranged the 2.0 kb Eco RI band containing the V_β8.2 gene segment; thus, at least two of the three functional members of the V_β8 subfamily are used in T cells. Using this antibody, 2% of human peripheral blood lymphocytes were found to express a member of the V_β8 subfamily in their β chains (32). The observation that the two (or three) functional V_β gene segments in this subfamily are used in such a large fraction of peripheral T cells implies that the human V_β gene segment repertoire is limited in size, as is the case in mouse (12–14). An analysis of a large number of human V_β gene segments also supports these data (15).

V_β Gene Segment Flanking Regions. Analyses of the DNA sequences of the flanking regions of the members of the V_β8 subfamily have identified several sequences that resemble eukaryotic gene transcriptional control sequences found adjacent to other genes (33). Eukaryotic promoter sequences generally include an A/T-rich region with a canonical ATA sequence called the TATA sequence located 21–23 bp 5′ to the initiation point of transcription, and the sequence CCAAT, which is found 5′ to the TATA sequence (33). In addition, immunoglobulin V gene segments have an additional region of transcription control located just 5′ to the TATA sequence, known as the octamer sequence (34, 35).

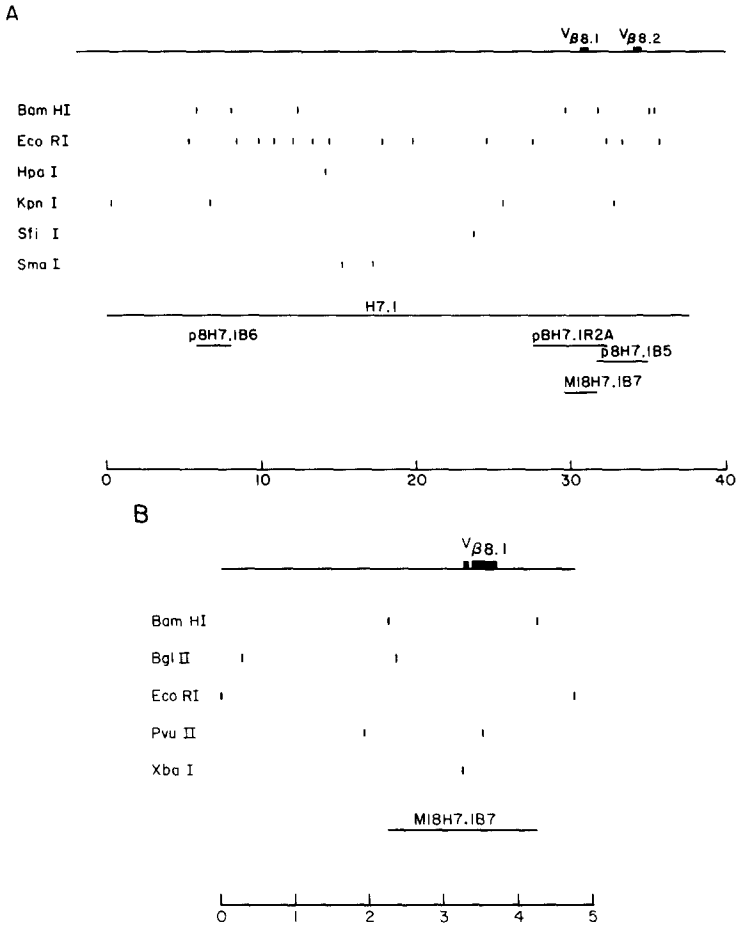

FIGURE 2. Restriction maps of the cosmid and λ clones containing the members of the $V_\beta 8$ family. (*A*) Cosmid H7.1, containing $V_\beta 8.1$ and $V_\beta 8.2$, and the pBR322 subclone pBH7.1R2A, the pUC8 subclones p8H7.1B6 and p8H7.1B5, and the M13mp18 subclone M18H7.1B7 of this region. (*B*) Plasmid pBH7.1R2A, containing the $V_\beta 8.1$ gene segment, and the M13mp18 subclone M18H7.1B7. (*C*) Cosmids H18.1 and H9.1, containing $V_\beta 8.4$, and the pBR322 subclone pBH9.1R3 and the pUC8 subclone p8H18.1RH that contains portions of these cosmids. (*D*) λVB8.5, containing the $V_\beta 8.5$ gene segment, and the M13mp18 subclone M18VB8.5 that contains a portion of this region. (*E*) λgt7 clone λVB8.3, containing the $V_\beta 8.3$ gene segment. Scale in each figure is in kb. The thin and thick boxes represent the first and second exons of the germline V_β gene segment.

The functional $V_\beta 8.1$ and $V_\beta 8.2$ gene segments have TATA-like sequences located both 25 and 89 bp 5′ to the methionine codon of the leader sequence. These gene segments also have a CCAAT-like sequence, TGGCCCATTC, located 34 bp 5′ to the second TATA-like sequence. The $V_\beta 8.3$ gene segment does not appear to have any of these promoter-like sequences. The only conserved sequence in the 5′ flanking region between $V_\beta 8.3$ and the other members of the $V_\beta 8$ subfamily is a 16 bp sequence located 5′ to the initiation codon of the leader (Fig. 3). It is possible that the 5′ untranslated region is much longer, or that an intron exists in this region for the $V_\beta 8.3$ gene segment. It is also possible

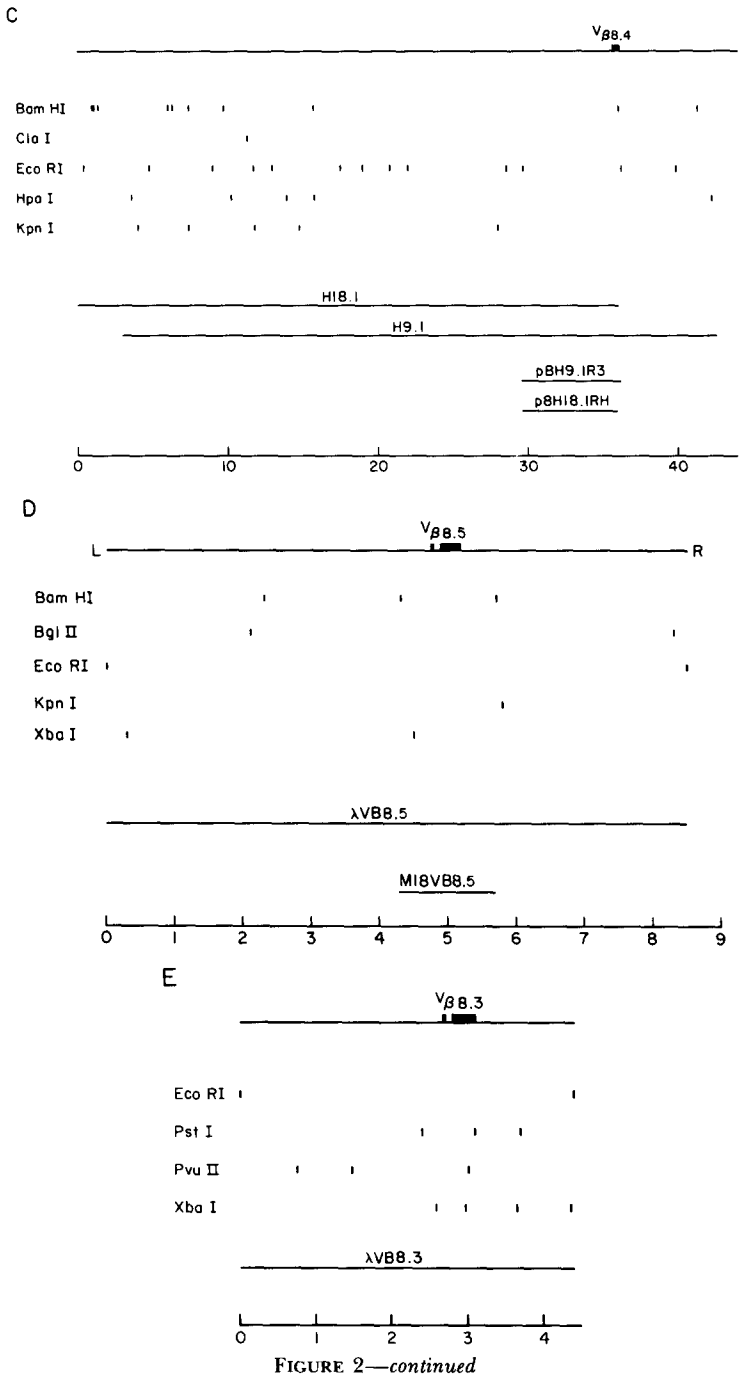

FIGURE 2—*continued*

that transcription of a β gene that uses the $V_\beta 8.3$ gene segment does not use the typical promoter sequences. Finally, it is possible that $V_\beta 8.3$ is a pseudogene due to a deletion of the promoter region; such a deletion is believed to have generated

FIGURE 3. Complete nucleotide sequence of the five members of the $V_\beta 8$ family and their flanking regions. Gaps were introduced to maximize homology. Leader and V coding regions are boxed and indicated. Putative TATA and CCAAT boxes in the 5′ region, termination codons for $V_\beta 8.4$ and $V_\beta 8.5$ in the variable gene segment sequence, and the rearrangement recognition signals in the 3′ flanking region are overlined and indicated. The 21 bp direct repeat in the 5′ flanking region of $V_\beta 8.5$ is indicated by two arrows, and the 16 bp conserved region in the 5′ flanking region of the $V_\beta 8$ subfamily members is indicated by a bracket.

a pseudogene in the T15 V_H gene subfamily (Siu, G., S. Crews, E. Springer, H. Huang, and L. Hood, manuscript in preparation). The V_β8.4 pseudogene appears to have the TATA-like sequence at position 89, and the CCAAT-like sequence described above. The V_β8.5 pseudogene gene segment does not appear to have a TATA-like sequence, although there is an A-rich sequence at position 89. The V_β8.5 gene segment does have the CCAAT-like sequence located 34 bp 5′ to the position 89 A/T-rich sequence. All of the V_β8 subfamily members appear to have a short (16 bp) conserved sequence in the 5′ flanking region; this sequence does not resemble any of the previously characterized promoter sequences, although its conservation between all of the members is puzzling in light of the lack of homology in this region in general. Although these data are provocative, additional data will be necessary to determine conclusively whether T cell receptor–specific promoter sequences exist.

V_β Gene Segment Evolution. One of the more intriguing aspects of the V_β gene segment family is the fact that most of the V_β subfamilies consist of only a few members. In mice, most of the subfamilies consist of only one member (12–14), and in humans, the largest subfamily consists of only six members (15). In contrast, the immunoglobulin V subfamilies range in size from 4 to >50 members (16–18). The reasons for this difference in subfamily size are unclear. One of the more attractive hypotheses that was proposed to explain this difference was that the V_β gene segments were mutating at a higher rate than the immunoglobulin V gene segments (36). This implies that the V_β segments are under relatively little selection pressure, even less than that of the immunoglobulin V gene segments, or alternatively, are under selection pressure to diverge. Evidence supporting this theory was obtained by hybridizing the mouse V_β probes to a wide variety of mammalian genomic DNAs. It was determined that most of the murine V_β probes could not detect V_β gene segments in even closely-related mammalian species using lower-stringency hybridizations and washes, implying that the mutation rate of the V_β gene segments was indeed very rapid (12).

It is possible to determine the mutation rate of V_β gene segments directly by using the method of Kimura (37). Using this method, the mutation frequency (the number of mutations per site, denoted K) and the mutation rate (number of mutations per site per year) between homologous sequences can be determined. The mutation frequency provides a method of comparing the extent to which two very similar genes are diverging from one another. The mutation rate permits the analysis of the rate at which a gene is mutating in units of mutations per year. The number of mutation events that have occurred between two similar sequences are counted and analyzed in two ways. In the first case, the mutations that resulted in replacements in the amino acid sequence (replacement-site mutations) and those that did not result in amino acid replacements (silent-site mutations) are categorized. In the second case, the mutations in the first, second, and third positions of each codon were separated. These techniques provide a method by which one can determine the nature of the mutation events that are occurring. The first method takes into consideration the fact that mutations that alter the primary amino acid sequence occur at different rates than those that do not. The second method takes into account differences in the mutation rates at different positions of the codon (38). The second position of the codon has no

TABLE II
Mutation Frequencies in the $V_\beta 8$ Family

Comparison	K_1*	K_2*	K_3*	K_A‡	K_S‡	K_A/K_S
$V_\beta 8.1$ vs. $V_\beta 8.2$	0.011 ± 0.01	0.011 ± 0.01	0.038 ± 0.02	0.0092 ± 0.006	0.049 ± 0.03	0.19
$V_\beta 8.1$ vs. $V_\beta 8.3$	0.22 ± 0.05	0.092 ± 0.03	0.29 ± 0.06	0.14 ± 0.03	0.47 ± 0.12	0.29
$V_\beta 8.1$ vs. $V_\beta 8.4$	0.25 ± 0.06	0.17 ± 0.05	0.33 ± 0.08	0.19 ± 0.03	0.50 ± 0.14	0.38
$V_\beta 8.1$ vs. $V_\beta 8.5$	0.30 ± 0.07	0.20 ± 0.05	0.39 ± 0.08	0.22 ± 0.04	0.55 ± 0.15	0.40
$V_\beta 8.2$ vs. $V_\beta 8.3$	0.24 ± 0.06	0.092 ± 0.03	0.30 ± 0.07	0.14 ± 0.03	0.50 ± 0.13	0.28
$V_\beta 8.2$ vs. $V_\beta 8.4$	0.28 ± 0.07	0.14 ± 0.04	0.33 ± 0.08	0.19 ± 0.03	0.52 ± 0.15	0.37
$V_\beta 8.2$ vs. $V_\beta 8.5$	0.30 ± 0.07	0.20 ± 0.05	0.39 ± 0.08	0.23 ± 0.04	0.56 ± 0.14	0.41
$V_\beta 8.3$ vs. $V_\beta 8.4$	0.33 ± 0.07	0.16 ± 0.05	0.36 ± 0.08	0.22 ± 0.04	0.57 ± 0.17	0.39
$V_\beta 8.3$ vs. $V_\beta 8.5$	0.36 ± 0.08	0.19 ± 0.05	0.49 ± 0.10	0.25 ± 0.04	0.77 ± 0.23	0.33
$V_\beta 8.4$ vs. $V_\beta 8.5$	0.32 ± 0.07	0.18 ± 0.05	0.45 ± 0.10	0.23 ± 0.04	0.71 ± 0.22	0.33
$V_\beta 11$ vs. $V_\beta 8.2$	0.25 ± 0.06	0.16 ± 0.05	0.43 ± 0.09	0.18 ± 0.03	0.75 ± 0.24	0.24

Error limits are 1 SD.
* K_1, K_2, and K_3 represent the mutation frequencies in the first, second, and third positions of the codon, respectively.
‡ K_A and K_S represent the mutation frequencies in the amino acid replacement- and silent-sites.

silent sites, and mutations in the second position are more likely to result in nonconservative replacements; that is, amino-acid replacement changes in the second position result in greater changes in the physical properties of the protein. Thus, mutation rates and frequencies in the second position for most genes are very low. In contrast, the third position of the codon has the most silent sites, and therefore the mutation rates and frequencies in the third position for most genes are relatively high. Finally, it is possible, using these data, to determine the divergence times between similar genes, using the value of the silent-site mutation rate as 5.1×10^{-9} mutations/site/yr (39).

Analysis of the mutation frequencies between the $V_\beta 8$ subfamily members show that the mutation frequencies in the replacement sites are much lower than the mutation frequencies in the silent sites, presumably reflecting selection pressure operating to maintain the coding region sequence. The ratios of the replacement-site mutation frequency to the silent-site mutation frequency (K_A/K_S) between the members of this subfamily range from 0.19 to 0.41, but from only 0.19 to 0.29 if comparisons of only the functional gene segments are considered (Table II). This is comparable to that observed in most eukaryotic genes (39), and much lower than that observed in immunoglobulin V_H gene subfamilies (Siu, G., et al., manuscript in preparation). In comparing the mutation frequencies in the three positions of the codon, it is apparent that the lowest mutation frequency is in the second position (K_2), and the highest mutation frequency is in the third position (K_3; Table II). This is similar to other eukaryotic genes and reflects the asymmetries in codon assignment described above (38). These data also contrast with the T15 V_H gene subfamily; in this case the lowest frequency is in the first position (K_1), and the frequency in the second position is consistently higher than that of the third position (Siu, G., et al., manuscript in preparation). These data imply that the V_β gene segments of the $V_\beta 8$ subfamily are mutating in the same manner as other eukaryotic genes and in a distinctly different manner than the immunoglobulin V gene segments.

A murine V_β gene segment homologous to the $V_\beta 8$ family was recently identified in a β chain cDNA isolated from a mouse spleen cDNA library (14). Unlike the human $V_\beta 8$ subfamily, the corresponding murine subfamily consists

of only one member. This gene segment, denoted $V_\beta 11$, is most closely related to the $V_\beta 8.2$ gene segment (Table I). Like the comparisons between the functional members of the $V_\beta 8$ subfamily in humans, comparisons between this murine V_β gene segment and the functional members of the human $V_\beta 8$ subfamily reveal that the K_A/K_S ratio and the relative values of K_1, K_2, and K_3 are similar to other eukaryotic genes, indicating that selection pressure is acting to maintain the coding sequence (Table II). Using the K_A and assuming that humans and mice diverged at mammalian radiation (85 million years ago [40]), the mutation rate in the amino acid replacement site is 10^{-9}/site·yr, which is comparable to the mutation rates in β globin genes, 1.13×10^{-9}/site·yr (41), and less than that of IFN-α, 2.1×10^{-9}/site·yr (42). Analyses of a large number of different eukaryotic genes have indicated that most have amino acid replacement-site mutation rate in the range of $0.2–2.0 \times 10^{-9}$/site·yrs (43). Thus, our data indicate that the V_β gene segments are mutating in a manner similar to that of other genes, and different from those of immunoglobulin V gene segments.

V_β Pseudogenes. Two of the five members of the $V_\beta 8$ subfamily, $V_\beta 8.4$ and $V_\beta 8.5$, are pseudogenes in that they share significant homology with a functional gene, but have mutations that would prevent their expression. Both have in-frame stop codons that would prevent translation of a variable region gene that uses one of these gene segments. The presence of two pseudogenes in a subfamily of five members indicates that the proportion of pseudogenes in the multimembered V_β subfamilies may be high, perhaps as high as that of the immunoglobulin V gene subfamilies (44). In addition, V_β pseudogenes have been identified in β chain cDNA clones isolated from murine thymus cDNA libraries (14). The high proportion of pseudogenes in the V gene subfamilies indicates that the generation of germline diversity may result in a large number of nonfunctional gene segments.

Chromosomal Organization of V_β Gene Segments. Characterization of the genomic clones containing the five members of the $V_\beta 8$ family indicate that two of the five members, $V_\beta 8.1$ and $V_\beta 8.2$, are only 3 kb apart. As mentioned above, these two gene segments are most homologous to each other, and are probably the result of a recent gene duplication event. The other members of the family do not appear to be closely linked. The $V_\beta 8.3$ gene segment was isolated on a 4.4 kb Eco RI fragment that does not overlap with any of the other clones. The $V_\beta 8.4$ gene segment was isolated on two overlapping cosmids containing 35 kb of 5′ flanking DNA and 6 kb of 3′ flanking DNA that are unlinked to any of the other clones. The $V_\beta 8.5$ gene segment was isolated on a 8.5 kb Eco RI fragment that does not overlap with any of the other clones. The $V_\beta 8.1$-$V_\beta 8.2$ cluster was isolated on one cosmid along with 29 kb of 5′ flanking DNA and 5 kb of 3′ flanking DNA. To characterize the genomic organization of the $V_\beta 8$ subfamily in greater detail, field-inversion gradient gel electrophoresis was used Lai and L. E. Hood, manuscript in preparation). This novel technique permits the separation and analysis of large fragments of DNA for gene linkage. The $V_\beta 8$-specific probe hybridizes to multiple bands of human DNA digested with the restriction enzymes Xho I, Cla I, and Sma I, ranging in size from <50 to >150 kb. However, a single 110 kb restriction fragment was detected when human DNA is digested with Sfi I (Fig. 4). Hybridization probes that were

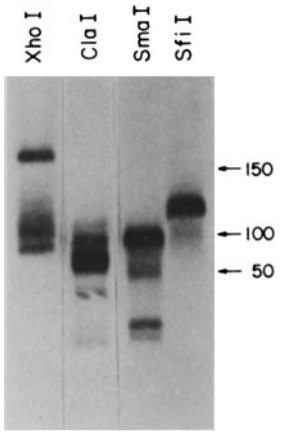

FIGURE 4. Field-inversion gel electrophoresis analyses of human HeLa cell DNA using the V_β8-specific hybridization probe from the YT35 cDNA. The restriction enzymes used to digest the DNA in each lane are indicated. Migration positions of the 150, 100, and 50 kb ligated λ DNA markers are indicated by arrows. Hybridization conditions are described in the Materials and Methods.

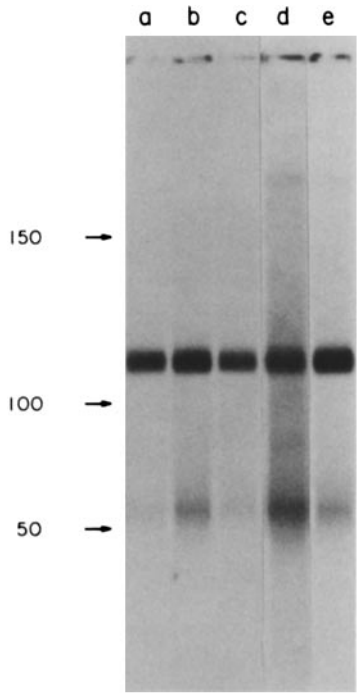

FIGURE 5. Hybridization of the V_β8 subfamily probes to HeLa DNA digested with Sfi I. Probes are derived from (*A*) V_β8.5, (*B*) V_β8.1, (*C*) V_β8.2, (*D*) V_β8.3, and (*E*) V_β8.4. Migration positions of the 150, 100, and 50 kb ligated λ DNA markers are indicated by arrows. Hybridization conditions are described in the Materials and Methods.

isolated from the cosmid and λ clones containing the members of the $V_\beta 8$ subfamily also hybridize to a single 110 kb Sfi I fragment (Fig. 5). We feel these data strongly indicate that the five V_β gene segments of the V_β subfamily are all present on a single 110 kb DNA fragment, although it is possible that the members of the $V_\beta 8$ subfamily hybridizes to similar-sized but different Sfi I fragments. Mapping of the λ and cosmid genomic clones containing the members of this subfamily have identified only one Sfi I site, located 6.6 kb 5′ to the $V_\beta 8.1$-$V_\beta 8.2$ cluster (Fig. 2*A*). This result is consistent with both models, although it indicates that, if the members of this subfamily are linked on the same Sfi I fragment, the $V_\beta 8.1$-$V_\beta 8.2$ cluster must be located 5′ to the other gene segments. The resolution of these two differing models will require the actual linkage of the members of this family using genomic clones.

If in fact the members of this subfamily are present on a single Sfi I fragment, this would be the first complete linkage of V subfamily members, and would indicate that, although the gene segments can be as close as 3 kb, the averaged spacing distance between subfamily members may be much larger. This is comparable to what is observed in the immunoglobulin V_H family, where the spacing distance between subfamily members can range from 3–5 kb (44, 45) to >23 kb (46 and S. Crews, E. Springer, G. Siu, and L. E. Hood, unpublished data). The available data do not rule out the possibility that gene segments that do not crosshybridize with the $V_\beta 8$ probe are interspersed between the subfamily members; such interspersion of nonhomologous V gene segments has been found in V_H gene segments (47). If this is the case for the V_β family, the spacing distances between gene segments will be correspondingly less.

Our characterization of the $V_\beta 8$ subfamily reveals that the structure and organization of the human V_β locus is similar to that of the immunoglobulin V gene loci. These similarities underscore the close structural and evolutionary relationships between immunoglobulins and the T cell antigen receptor.

Summary

We have isolated and sequenced five germline V_β gene segments that are homologous to the V region of the YT35 cDNA encoding the β chain of the T cell antigen receptor from the tumor MOLT-3. One of these gene segments is identical to the YT35 V segment, and therefore is the corresponding germline V_β gene segment encoding the YT35 cDNA. The other four V_β members exhibit 77–98% homology to the YT35 V gene segment. Two of these V_β gene segments are pseudogenes. Analyses of the coding region sequences reveal that, although the V_β gene segments are very diverse, they are mutating at a rate comparable to that observed in most eukaryotic genes. Analyses of the genomic clones show that the spacing distance between germline V_β gene segments ranges from 3 kb to >30 kb, and the entire $V_\beta 8$ subfamily appears to be linked by a total of no more than 110 kb of DNA.

We thank Dr. Suzanna Horvath and Marilyn Tomich for providing synthetic oligonucleotide primers, Mr. Lance Fors for providing the human cosmid library, and Debora Maloney for excellent technical assistance. We also thank Drs. Richard Barth, Patrick Concannon, Joan Kobori, Mitchell Kronenberg, and Mr. Tim Hunkapiller for critique of the manuscript, and Connie Katz, Susan Mangrum and the other members of the

secretarial staff at the Division of Biology at the California Institute of Technology for the preparation of the manuscript.

Received for publication 27 June 1986.

References

1. Davis, M. M., Y.-H. Chien, N. R. J. Gascoigne, and S. M. Hedrick. 1984. A murine T-cell receptor gene complex: Isolation, structure and rearrangement. *Immunol. Rev.* 81:235.
2. Kronenberg, M., G. Siu, L. Hood, and N. Shastri. 1986. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Ann. Rev. Immunol.* 4:529.
3. Yanagi, Y., Y. Yoshikai, K. Leggett, S. P. Clark, I. Aleksander, and T. W. Mak. 1984. A human T-cell specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* (*Lond.*). 308:145.
4. Hedrick, S. M., E. A. Nielsen, J. Kavaler, D. I. Cohen, and M. M. Davis. 1984. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature* (*Lond.*). 308:153.
5. Chien, Y.-H., N. R. J. Gascoigne, J. Kavaler, N. E. Lee, and M. M. Davis. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* (*Lond.*). 309:322.
6. Siu, G., S. Clark, Y. Yoshikai, M. Malissen, Y. Yanagi, E. Strauss, T. Mak, and L. Hood. 1984. The human T-cell antigen receptor is encoded by variable, diversity and joining gene segments that rearrange to generate a complete V gene. *Cell.* 37:393.
7. Clark, S. P., Y. Yoshikai, S. Taylor, G. Siu, L. Hood, and T. W. Mak. 1984. Identification of a diversity segment of the human T-cell receptor beta chain, and comparison to the analogous murine element. *Nature* (*Lond.*). 311:387.
8. Kavaler, J., M. M. Davis, and Y.-H. Chien. 1984. Localization of a T-cell receptor diversity-region element. *Nature* (*Lond.*). 310:421.
9. Siu, G., M. Kronenberg, E. Strauss, R. Haars, T. Mak, and L. Hood. 1984. The structure, rearrangement and expression of D_β gene segments of the murine T-cell antigen receptor. *Nature* (*Lond.*). 311:344.
10. Gascoigne, N. R. J., Y.-H. Chien, D. M. Becker, J. Kavaler, and M. M. Davis. 1984. Genomic organization and sequence of T-cell receptor β-chain constant- and joining-region genes. *Nature* (*Lond.*). 310:387.
11. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Governman, T. Hunkapiller, M. Prystowsky, Y. Yoshikai, F. Fitch, T. 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.
12. Patten, P., T. Yokota, J. Rothbard, Y.-H. Chien, K.-I. Arai, and M. Davis. 1984. Structure, expression and diversity of T-cell receptor β-chain variable regions. *Nature* (*Lond.*). 312:40.
13. Barth, R., B. Kim, N. Lan, T. Hunkapiller, N. Sobieck, A. Winoto, H. Gershenfeld, C. Okada, D. Hansburg, I. Weissman, and L. Hood. 1985. The murine T-cell receptor employs a limited repertoire of expressed V_β gene segments. *Nature* (*Lond.*). 316:517.
14. Behlke, M. A., D. G. Spinella, H. S. Chou, W. Sha, D. L. Hartl, and D. Y. Loh. 1985. T-cell receptor β chain expression: Dependence on relatively few variable region genes. *Science* (*Wash. DC.*). 229:566.
15. Concannon, P., L. A. Pickering, P. Kung, and L. E. Hood. 1986. Diversity and structure of human T-cell receptor V_β genes. *Proc. Natl. Acad. Sci. USA.* In press.
16. Cory, S., B. Tyler, and J. Adams. 1981. Sets of immunoglobulin V_κ genes homologous

to ten cloned V_κ sequences: Implications for the number of germline V_κ genes. *J. Mol. Appl. Genet.* 1:103.

17. Brodeur, P., and R. Riblet. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol.* 14:922.
18. Dildrop, R. 1984. A new classification of mouse V_H sequences. *Immunol. Today.* 5:85.
19. Arden, B., J. L. Klotz, G. Siu, and L. E. Hood. 1985. Diversity and structure of genes of the α family of mouse T-cell antigen receptor. *Nature (Lond.).* 316:783.
20. Becker, D., P. Patten, Y.-H. Chien, T. Yokota, Z. Eshkar, M. Giedin, N. R. J. Gascoigne, C. Goodenow, R. Wolf, K.-I. Arai, and M. Davis. 1985. Variability and repertoire size in T-cell receptor V_α gene segments. *Nature (Lond.).* 317:430.
21. Carle, G. F., M. Frank, and M. V. Olson. 1986. Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science (Wash. DC).* 232:65.
22. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.
23. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
24. 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 I. *J. Mol. Biol.* 113:237–251.
25. Kobori, J. A., A. Winoto, J. McNicholas, and L. Hood. 1984. Molecular characterization of the recombination region of six murine major histocompatibility complex (MHC) I-region recombinants. *J. Mol. Cell Immunol.* 1:125–131.
26. Smith, H. O., and M. L. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucl. Acids Res.* 3:2387.
27. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.
28. Strauss, E., J. Kobori, G. Siu, and L. Hood. 1986. Specific-primer–directed DNA sequencing. *Anal. Biochem.* 154:353–360.
29. Sims, J. E., A. Tunnacliffe, W. J. Smith, and T. H. Rabbitts. 1984. Complexity of human T-cell antigen receptor β-chain constant- and variable-region genes. *Nature (Lond.).* 312:541.
30. Goverman, J., K. Minard, N. Shastri, T. Hunkapiller, D. Hansburg, E. Sercarz, and L. Hood. 1985. Rearranged β T-cell receptor genes in a helper T-cell clone specific for lysozyme: No correlation between V_β and MHC restriction. *Cell.* 40:859.
31. Ikuta, K., T. Ogura, A. Shimizu, and T. Honjo. 1985. Low frequency of somatic mutation in β chain variable region genes for human T-cell receptors. *Proc. Natl. Acad. Sci. USA.* 82:7701.
32. Acuto, O., T. J. Campen, H. D. Royer, R. E. Hussey, C. B. Poole, and E. L. Reinherz. 1985. Molecular analysis of T-cell receptor (Ti) variable region (V) gene expression. Evidence that a single Ti β V gene family can be used in formation of V domains on phenotypically and functionally diverse T-cell populations. *J. Exp. Med.* 161:1326.
33. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50:349.
34. Falkner, F. G., and H. G. Zachau. 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature (Lond.).* 310:71.
35. 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.

36. Davis, M. M. 1985. Molecular genetics of the T-cell receptor beta chain. *Ann. Rev. Immunol.* 3:537.
37. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* 78:454.
38. Salser, W. 1976. Globin mRNA sequences: Analysis of base pairing and evolutionary implications. *Cold Spring Harbor Symp. Quant. Biol.* 62:985.
39. Miyata, T., T. Yasunaga, and T. Nishida. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proc. Natl. Acad. Sci. USA.* 77:7328.
40. Romero-Herrera, A. E., H. Lehman, K. A. Joysey, and A. E. Friday. 1973. Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. *Nature (Lond.).* 246:389.
41. Miyata, T., H. Hayashida, R. Kikuno, M. Hasegawa, M. Kobayashi, and K. Koike. 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* 19:28.
42. Miyata, T., and H. Hayashida. 1982. Recent divergence from a common ancestor of human IFN-α genes. *Nature (Lond.).* 295:165.
43. Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2:150.
44. Givol, D., R. Zakut, K. Effron, G. Rechavi, D. Ram, and J. B. Cohen. 1981. Diversity of germ-line immunoglobulin V_H genes. *Nature (Lond.).* 292:426.
45. Bothwell, A. L. M., M. Poskind, M. Reth, T. Imaniski-Kari, K. Rajewsky, and D. Baltimore. 1981. Heavy chain variable region contribution to the NP^a family of antibodies: Somatic mutation evident in a γ2A variable region. *Cell.* 24:625.
46. Kemp, D., B. Tyler, O. Bernard, N. Gough, S. Gerondakis, J. M. Adams, and S. Cory. 1981. Organization of genes and spacers within the mouse immunoglobulin V_H locus. *J. Mol. Appl. Genet.* 1:245.
47. Johnson, M. J., A. M. Natali, H. M. Cann, T. Honjo, and L. L. Cavalli-Storza. 1984. Polymorphisms of a human variable heavy chain gene show linkage to constant region genes. *Proc. Natl. Acad. Sci. USA.* 81:7840.