

# The $V_{\beta}17^+$ T Cell Repertoire: Skewed $J\beta$ Usage after Thymic Selection; Dissimilar CDR3s in $CD4^+$ Versus $CD8^+$ Cells

By Serge Candéas, Caroline Waltzinger, Christophe Benoist, and Diane Mathis

*From the Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 67085 Strasbourg, France*

## Summary

To ascertain how the actual repertoire of T cell receptors (TCRs) deviates from the theoretical, we have generated a large number of junctional region sequences from TCRs carrying the  $V_{\beta}17$  variable region. The >600 sequences analyzed represent transcripts from nine different cell populations, permitting several comparisons: transcripts from an expressed vs. a non-expressed  $V_{\beta}17$  allele, those from  $E^+$  vs.  $E^-$  mice, transcripts from immature vs. mature thymocytes, those from thymic vs. peripheral T cells, and those from  $CD4^+$  vs.  $CD8^+$  cells. These comparisons have allowed us to distinguish between the influence of molecular events involved in TCR gene rearrangement and that of various selection events that shape the T cell repertoire. Our most striking findings are: (a) that  $J\beta$  usage is markedly skewed, partly due to recombination mechanics and partly due to selection forces: in particular, those mediated by the class II E molecule in the thymus; and (b) that TCRs on  $CD4^+$  and  $CD8^+$  cells show intriguing dissimilarities. In addition, we present evidence that N nucleotide additions occur with clear biases, probably due to idiosyncrasies of the recombination enzymes, and provide arguments that TCR and immunoglobulin CDR3s have distinct structures.

Theoretically, the murine immune system can make use of an enormous repertoire of  $\alpha/\beta$  T cells. Any rearranged TCR  $\beta$  chain gene is created by the juxtaposition of 1 of 22 V, 1 of 2 D, and 1 of 12 J region segments (for reviews, see references 1 and 2). At each of the juxtaposed segment termini (3'V, 5'D, 3'D, 5'J) one or two so-called P nucleotides can be added, almost certainly as byproducts of the mechanics of rearrangement (3); some of the terminal nucleotides can then be nibbled away by an undefined exonuclease activity (1, 2). Finally, a variable number of so-called N nucleotides can be inserted at the V-D and D-J junctions by a template-independent mechanism (1, 2). Assuming that these processes operate randomly, and considering similar processes to be in play with the  $\alpha$  chain genes, Davis and Bjorkman (4) have calculated that there may be as many as  $10^{15}$   $\alpha/\beta$  TCR specificities available to the mouse.

But is this enormous repertoire actually realized? Already there have been some indications that it is not. Some mice lack particular V (5, 6) or J (7) gene segments, or do not express certain ones because of mutations (8). Some animals delete from the peripheral repertoire almost all T cells displaying particular  $V_{\beta}$ s, seemingly because they are self reactive (for reviews, see references 9 and 10). In this report, we

have sought to determine whether the junctional regions of TCR  $\beta$  chains really have a haphazard constitution. According to Davis and Bjorkman (4), the TCR junctional regions should make by far the greatest contribution to their diversity. We have asked questions like: is J region usage entirely random? Is D region usage? Are P nucleotides found at random? What are the constraints on exonucleolytic nibbling? Are N nucleotides really added randomly? We have posed these questions in a variety of contexts chosen to permit a distinction between the various forces potentially capable of skewing the repertoire, e.g., the molecular events involved in TCR gene rearrangement, selection events during E-mediated clonal deletion in the thymus, and selection events predicated MHC class I vs. class II restriction.

The extensive data we have accumulated provide a detailed picture of the actual  $V_{\beta}17^+$  TCR repertoire, and provide some hint of the forces that shape it, some previously unrecognized.

## Materials and Methods

*Mice.* C57 Bl/6J (B6) and SJL/J (SJL) mice were obtained from Iffa-Credo (Les Oncins, France) and The Jackson Laboratory (Bar

Harbor, ME), respectively. The SJL.E $\alpha$  line was created by repeatedly backcrossing the E $\alpha$ 16 line onto SJL; for the experiments presented here, mice in the seventh or eighth backcross generation were used. The E $\alpha$ 16 line has been described (11–13); it carries an E $\alpha$ <sup>k</sup> transgene that is expressed entirely normally.

**T Cells and RNA.** Thymocytes or lymph node cells were prepared from individual mice, and were incubated with a cocktail of antibodies: anti-V $\beta$ 17 (KJ23; 14) followed by Texas Red-tagged anti-mouse IgG, PE-conjugated anti-CD4, and fluorescein-conjugated anti-CD8, as described (11, 13). The triple-labeled cells were sorted on a cytofluorimeter (ATC 3000; ODAM, Wissembourg, France) equipped with dual-laser excitation into CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17<sup>+</sup>, CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17<sup>+</sup>, CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17<sup>+</sup>, or CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17<sup>+</sup> populations of from  $3 \times 10^3$  to  $10^5$  cells.  $1$  or  $2 \times 10^6$  HeLa cells were added to each sample as carrier. RNA was prepared by the standard NP-40 lysis method (15) and was stored in water at  $-20^\circ\text{C}$ .

**PCR Amplifications.** cDNA was synthesized from 0.5–1  $\mu\text{g}$  of RNA by the action of avian myeloblastosis virus reverse transcriptase in the buffer recommended by Cetus Corp. (Emeryville, CA) for PCR incubations. cDNA/RNA duplexes were denatured for 5 min at  $95$ – $100^\circ\text{C}$ . After quick-cooling on ice, the mixture was supplemented with Taq DNA polymerase (1 U) and primers (50 pm), and the volume was brought up to 50  $\mu\text{l}$  with PCR buffer (Cetus Corp.). For amplification of V $\beta$ 17 sequences, the primers were 5'-CTTTGTCTGCTTGGGG-3' in the V region and 5'-AGCACACGAGGGTAGCCTT-3' in the C region. 30 cycles of amplification were performed in a "DNA Thermal Cycler" (Perkin Elmer/Cetus, Emeryville, CA). Each cycle consisted of 30 s at  $92^\circ\text{C}$ , 30 s at  $50^\circ\text{C}$ , and 1 min at  $72^\circ\text{C}$ . The last cycle was followed by an incubation at  $72^\circ\text{C}$  for 10 min. A second round of amplification was performed starting with 1  $\mu\text{l}$  of the first-round mixture. The PCR conditions were identical except that the V region primer was 5'-GTCCTGTGGCGCTGCAGGCTCTTTATGTTGCT-3' and the C region primer was 5'-GACAGAACTTTGAATTCCTCTGCTTTTGTATGG-3'. Both of these oligonucleotides contain artificial cloning sites (underlined): PstI in the former case, EcoRI in the latter.

**Cloning and Sequencing.** After the last amplification, the Taq polymerase was eliminated by phenol/chloroform then chloroform extraction, and the DNA was precipitated in ethanol. The pelleted material was redissolved and digested with PstI and EcoRI and then migrated on a 2.5% agarose gel. The band at  $\sim 450$  bp was excised and the DNA recovered by electroelution followed by precipitation in ethanol. The purified fragments were ligated with a vector prepared by digesting M13mp19 DNA with PstI and EcoRI. The ligated material was transfected into JM103 cells, and the resulting plaques were screened with either the oligo 5'-ACAGAGGTACAGTG-3' (V $\beta$ 17a) or the oligo 5'-ACAGAGCTACAGTG-3' (V $\beta$ 17b). Single-stranded DNA was prepared from positively hybridizing plaques and was sequenced by the dideoxy method using standard techniques.

**Screening for J $\beta$ 1.1.** Large numbers of M13 clones emanating from different T cell populations were doubly screened for V $\beta$ 17 and J $\beta$ 1.1. For the former, we used the PCR primer detailed above; for the latter, we used the oligo 5'-CTACAAGTGTGAGTGTGG-3'. Hybridization was at  $45^\circ\text{C}$  in  $2\times$  SSC, and washing was performed at  $37^\circ$  in  $2\times$  SSC.

**Contamination.** Elaborate precautions were taken to prevent sample contamination, a major problem of PCR-based techniques. All solutions were aliquoted, and aliquots were used only once. Aside from the customary negative controls, a mock sample was processed along with each set of experimental samples: droplets

of PBS without any cells were sorted and the entire procedure continued, including the screening of M13 plaques. This control ruled out contamination at any step along the way.

## Results

### Strategy

To evaluate how closely the actual T cell repertoire approximates the theoretical, we have chosen to focus on a single TCR V $\beta$ . This vastly reduces the number of sequences that must be generated to ensure statistical significance and greatly simplifies the sequence analysis. We have elected to concentrate on V $\beta$ 17 for a variety of reasons.

(a) V $\beta$ 17 is a particularly well-characterized TCR variable region, at both the gene and protein levels. cDNA and genomic sequences have been determined (14, 16, and K. Signorelli, unpublished results) and an effective anti-V $\beta$ 17 reagent is available (14).

(b) V $\beta$ 17 alleles are not always expressed as cell surface protein. A few inbred mouse strains (e.g., SJL) carry the expressed allele, V $\beta$ 17a1. Many others (e.g., B6) bear the nonexpressed counterpart, V $\beta$ 17b, which has a point mutation that specifies a stop codon, resulting in translation termination in the variable region (8). Thus, this allele is detectably transcribed into mRNA, but there is no corresponding protein at the cell surface. Hence, V $\beta$ 17b transcripts represent a virgin repertoire, incapable of being selected.

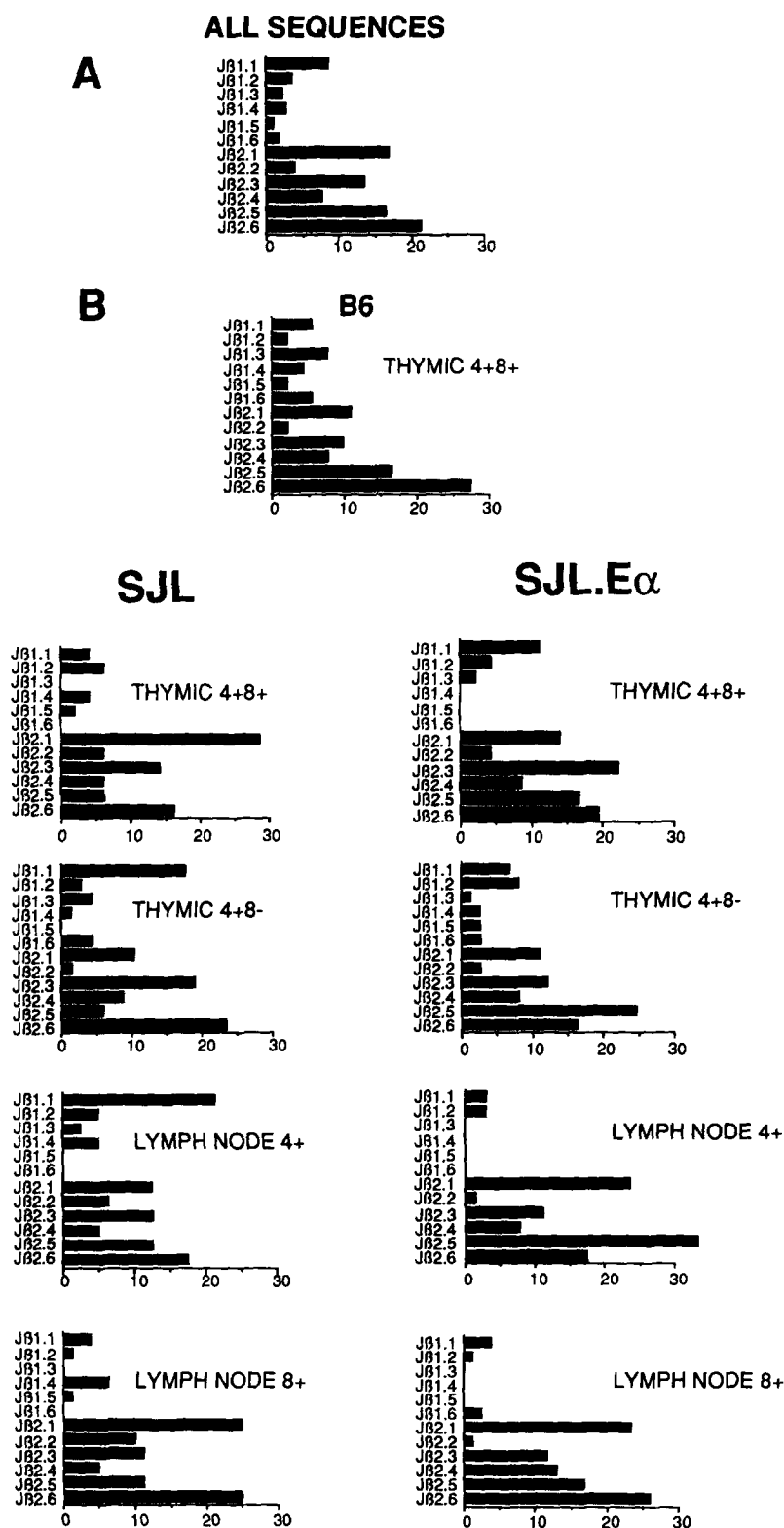
(c) V $\beta$ 17a<sup>+</sup> T cells have been observed to undergo both negative (17) and positive (18–20) selection. For example, clonal deletion of many, though not all, V $\beta$ 17a<sup>+</sup> T cells occurs within the thymus of mice that express the E complex and an as yet uncharacterized ligand. By repeatedly backcrossing an E $\alpha$  transgene (11) onto the SJL background, we have created a well-controlled system for studying this phenomenon: SJL vs. SJL.E $\alpha$ , strains that should differ only by expression of the E complex.

Our strategy relies on the power of PCR technology. Thymus or lymph node cells are divided into discrete populations by electronic sorting after staining with anti-CD4, -CD8, and -V $\beta$ 17 reagents. V $\beta$ 17 transcripts are amplified from the various pools by PCR using C $\beta$ - and V $\beta$ 17-specific oligos. And finally, 50–100 transcripts are sequenced for each population. Although we have routinely tried to work with  $\sim 10,000$  sorted cells, the technique is so sensitive that one can use  $<100$ .

Following this strategy, we have generated sequences from the following populations: thymus CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17b, thymus CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> and CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> from E<sup>−</sup> mice; lymph node CD4<sup>+</sup>V $\beta$ 17a<sup>+</sup> and CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> from E<sup>−</sup> mice; thymus CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> and CD4<sup>+</sup>CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> from E<sup>+</sup> mice; and lymph node CD4<sup>+</sup>V $\beta$ 17a<sup>+</sup> and CD8<sup>+</sup>V $\beta$ 17a<sup>+</sup> from E<sup>+</sup> mice. It should be kept in mind that each set of sequences routinely derives from three or more individual mice in order to ensure generality of the observations.

### Nucleotide Sequences

Due to an editorial decision, we do not present nucleotide



**Figure 1.** Frequency of individual J $\beta$  region usage. J $\beta$  region frequencies were calculated from the data sets, and are shown here as histograms. (A) All V $\beta$ 17a sequences are compiled. (B) individual histograms for each of the sorted thymic or lymph node populations from B6, SJL, or SJL.E $\alpha$  mice.

sequences for the V $\beta$ 17<sup>+</sup> TCRs from the nine cell populations. These can be obtained in raw form from the EMBL and GenBank data bases and in a considerably more useful annotated form directly from the authors. Several points emerge from our analysis of the data.

**J Region Usage.** A rearranged TCR  $\beta$  gene can use any one of 12 J segments, half associated with C $\beta$ 1 and the other half with C $\beta$ 2 (1, 2). Fig. 1 presents histograms of J region usage, both globally (A) and for individual sorted populations (B).

Considering the global histogram, one is immediately struck by the fact that J region usage is nonrandom. J $\beta$ 1.5 is used in only 0.9% of the TCRs, while J $\beta$ 2.6 is used 21.3% of the time. In general, there is a marked preference for J $\beta$ 2 over J $\beta$ 1: 80% vs. 20%. This skewing appears to be unrelated to thymic selection events because it occurs with both immature (thymic CD4<sup>+</sup>CD8<sup>+</sup>) and mature (peripheral CD4<sup>+</sup> and CD8<sup>+</sup>) populations, with both E<sup>+</sup> and E<sup>-</sup> mice, and even with the nonexpressed V $\beta$ 17b allele.

Considering the histograms for the different cell populations, one does see some evidence for selection of particular J $\beta$ s. For example, J $\beta$ 1.1 is highly enriched during the transition from CD4<sup>+</sup>CD8<sup>+</sup> to CD4<sup>+</sup>CD8<sup>-</sup> thymocytes in E<sup>-</sup> mice, increasing from 4.1% to 17.6%, and this heavy usage persists in peripheral CD4<sup>+</sup> cells, at a level of 21.3%. Interestingly, enrichment for J $\beta$ 1.1 is not observed in the thymic CD4<sup>+</sup>CD8<sup>-</sup> or the peripheral CD4<sup>+</sup> population of E<sup>+</sup> animals, nor in CD8 single-positives from either E<sup>+</sup> or E<sup>-</sup> mice. To fortify the conclusion that T cells expressing J $\beta$ 1.1 undergo selection, we screened large numbers of M13 plaques derived from the different populations with both a V $\beta$ 17 and a J $\beta$ 1.1 probe. As is evident from Table 1, this experiment produced results in complete accord with the sequencing data: there is a clear difference in the number of V $\beta$ 17a<sup>+</sup> clones that carry J $\beta$ 1.1 in the CD4<sup>+</sup> population of E<sup>+</sup> vs. E<sup>-</sup> mice, as well as in the CD4<sup>+</sup> vs. CD8<sup>+</sup> populations of both E<sup>+</sup> and E<sup>-</sup> animals.

J $\beta$ 2.5 usage also shows evidence of selection. In E<sup>+</sup> mice, passage from the CD4<sup>+</sup>CD8<sup>+</sup> to the CD4<sup>+</sup>CD8<sup>-</sup> thymus compartment is accompanied by an increase in frequency from 16.2 to 24.7%. This is further augmented in the periphery where nearly one-third of CD4<sup>+</sup> cells use this J $\beta$ . In E<sup>-</sup> mice, there is no such increase in either thymic or peripheral CD4 single-positives.

Both of these variations in J $\beta$  usage are statistically significant when assessed by the  $\chi^2$  method ( $p < 0.05$ ).

**D Region Usage.** The TCR  $\beta$  locus has two D segments, one associated with C $\beta$ 1 and the other with C $\beta$ 2 (1, 2). In calculating the frequency of D $\beta$ 1 vs. D $\beta$ 2 usage, we have only taken into account those sequences for which an unambiguous assignment can be made: not always possible given the marked sequence homology between the two segments,

**Table 1.** J $\beta$ 1.1 Usage in V $\beta$ 17<sup>+</sup> T Cells

Mouse strain	CD4	CD8
E negative	165/1,232 (13.4%)	43/1,024 (4.1%)
E positive	1/972 (0.1%)	66/1,140 (5.7%)

Values represent the numbers (and percentages) of M13 plaques that were positive after hybridization to a specific J $\beta$ 1.1 probe (see Materials and Methods for details). Plaques are derived from lymph node CD4<sup>+</sup> or CD8<sup>+</sup> T cells from SJL or SJL.E $\alpha$ .

coupled with the fact that extensive exonuclease digestion can occur during the rearrangement process. D $\beta$ 1 is preferred over D $\beta$ 2 both globally (54% vs. 46%) and in the unselectable V $\beta$ 17b cell population (62% vs. 37%). Skewing in the individual T cell populations is variable, with no obvious logic.

D $\beta$ 1 is found in combination with J $\beta$ 2 about twice as often as with J $\beta$ 1. This is true for all T cell populations.

**P Nucleotide Addition.** If no exonucleolytic nibbling has taken place, one or two P nucleotides (3) may be found at each of the four recombined segment termini: the 3' end of the V region, the 5' and 3' ends of the D segments, and the 5' end of the J segment.

Indeed, we find P nucleotides in a number of sequences. Interestingly, they are found more often at VD joints (at 67% of undigested 3'V or 5'D termini) than at DJ joints (45% of undigested 3'D or 5'J termini).

**Exonucleolytic Chewing.** Rearrangement of TCR and Ig genes often entails a variable degree of exonucleolytic nibbling of the recombining termini. Table 2 lists the number of nucleotides removed from the 3'V and 5'J termini (a similar analysis for the D termini would not be meaningful because the deletion assignments are often ambiguous).

At the 3'V end, the vast majority of sequences show removal of zero to four bases, but in a few cases the nibbling does extend further, up to 11 bases. A very similar profile is observed with the unexpressed V $\beta$ 17b sequences. At the 5'J end, removal of zero to four bases is also standard, but there are some surprising differences with the different J segments.

**Table 2.** Exonucleolytic Nibbling during V $\beta$ 17 TCR Rearrangement

No. of bases removed	3' end of V region		5' end of J region (V $\beta$ 17a sequences)		
	V $\beta$ 17a	V $\beta$ 17b	J $\beta$ 2.1	J $\beta$ 2.3	J $\beta$ 2.6
0	13.6	11.5	8.8(1)*	46.0	19.7(3)
1	14.6	9.6	23.5(2)	11.8(1)	5.3(0)
2	15.0	17.3	8.8	14.4	3.8(2)
3	17.9	15.3	10.8(1)	9.2(1)	12.1(5)
4	20.7	19.2	11.8(2)	5.2(1)	19.7(7)
5	5.3	7.6	11.8	6.5	13.6(1)
6	8.6	13.4	11.8	3.9	12.8(0)
7	1.8	1.9	5.9	1.3	11.4(2)
8	1.4	3.8	2.9	0	0.7(0)
9	0.3	0	2.9	1.3	0.7(0)
10	0	0	0	0	0 (0)
11	0.3	0	1.0	0	0 (0)
$n = 279 \quad n = 52 \quad n = 102 \quad n = 76 \quad n = 132$					

\* The values in parentheses indicate the corresponding number of occurrences in V $\beta$ 17b sequences.

For J $\beta$ 2.3, almost half of the sequences show no deletion. For J $\beta$ 2.6, the distribution seems bimodal: several sequences are undeleted, but the majority have three to six bases removed. The J $\beta$ 2.1 pattern seems different again. While the number of V $\beta$ 17b sequences is too small for a rigorous comparison, one notes that the V $\beta$ 17b percentages for J $\beta$ 2.6 seem to parallel the V $\beta$ 17a values. Thus, the peculiar distribution that we observe directly reflects molecular recombination events rather than selective influences, and exonucleolytic nibbling seems a fairly controlled process, sensitive to local DNA structure.

**N Nucleotide Addition.** A variable number of N nucleotides can be inserted at the VD and DJ junctions. For the V $\beta$ 17<sup>+</sup> TCRs analyzed here, the number added per sequence ranges from 0 to 14, reaching a maximum of 11 for a single junction. The average number of N nucleotides per sequence shows little variation in the different T cell populations, hovering between 2.2 and 3.4, with an average of 2.8.

As a further refinement, we sought to determine whether N region addition shows any nucleotide preferences. Table 3 lists the frequency with which each of the four bases is used at each junction, globally and in individual cell populations. Interestingly, base utilization at the VD and DJ junctions is not identical. The nucleotides inserted between V and D are rich in cytosine, while those added between D and J show no such enrichment: 38.9–57.9% vs. 10.6–23.3%.

### Protein Sequences

Protein sequences for the V $\beta$ 17<sup>+</sup> TCRs from the nine cell populations are presented in Fig. 2. Only the junctional regions are shown. Several points emerge as salient.

**CRD3 Lengths.** By analogy with Ig heavy chains, the CDR3 of a TCR  $\beta$  chain should extend from the serine situated at position 94 to the phenylalanine in the motif FGXG, conserved in almost all TCR J regions (21). Fig. 3 shows

histograms of CDR3 lengths, again globally (A) and for the individual sorted populations (B).

The overall average length is 10.0 amino acids, ranging from 6 to 15. The profiles from E<sup>-</sup> mice exhibit very little variability, while those from E<sup>+</sup> animals do show one difference worth mentioning: the CDR3s from peripheral CD8<sup>+</sup> cells are noticeably longer than those from peripheral CD4<sup>+</sup>s (10.7 amino acids on average, vs. 9.3).

**NDN Amino Acids.** The impressive variability of the TCR  $\beta$  chain CDR3 regions rests partly on N addition and partly on the D segment choice and its "processing". The particular residues that one finds in the NDN stretch are heavily influenced by the coding potential of the D segments. The amino acids that can be specified by D $\beta$ 1 and D $\beta$ 2 are indicated in Figure 4. It is worth keeping in mind that all three reading frames can be used and that glycine predominates in each.

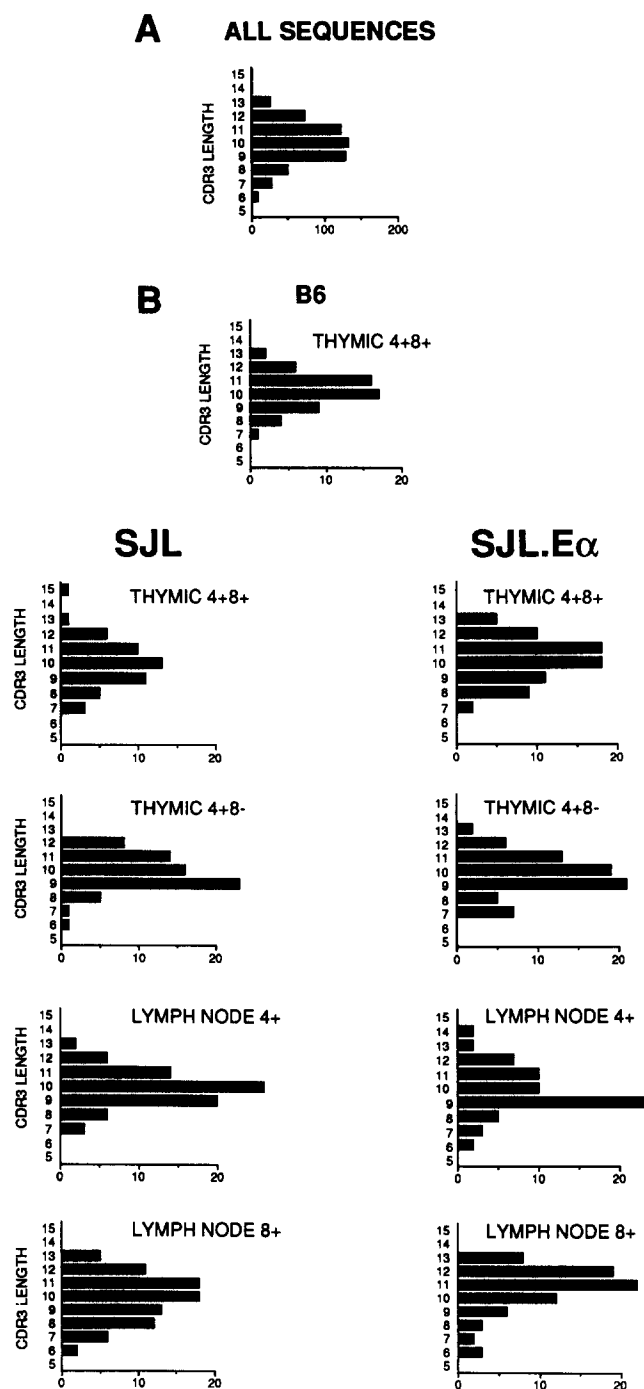
Table 4 illustrates amino acid usage in the NDN region of our V $\beta$ 17<sup>+</sup> TCRs. Clearly, the patterns are nonrandom. Considering first those amino acids encoded within D $\beta$ , we can make two points of interest. First, as expected, glycine is by far the most frequently used residue (fully one third of all the NDN amino acids are glycine). Second, one residue appears to be differentially used in the various cell populations (tryptophan is used distinctly more frequently in peripheral CD8<sup>+</sup> cells than in CD4<sup>+</sup> cells). Considering those amino acids not encoded within D $\beta$ , we can again bring out two points. First, cysteine is almost never found, perhaps not surprisingly considering its potential to interfere with the Ig fold. Second, proline is frequently observed, but this can probably be explained by the fact that its codon is CCN. As discussed above, N nucleotides inserted at the VD junction are enriched in Cs; in addition, P nucleotides are frequently detected at the 5'D terminus, where they occur as either one or two Cs.

**Table 3.** Base Distribution in N Nucleotides

	N nucleotides between V and D				N nucleotides between D and J				
	G	A	T	C	G	A	T	C	
	19.6	17.0	15.6	48.1	31.2	28.1	23.8	16.7	All sequences
V $\beta$ 17b,CD4 <sup>+</sup> 8 <sup>+</sup>	24.3	9.7	11.3	54.3	23.0	33.8	21.6	21.6	B6
Thy CD4 <sup>+</sup> 8 <sup>+</sup>	26.6	15.0	8.4	50.0	25.0	28.3	23.3	23.3	
Thy CD4 <sup>+</sup> 8 <sup>-</sup>	20.6	18.6	19.6	43.3	26.9	27.9	34.6	10.6	
LN CD4 <sup>+</sup>	8.0	19.3	14.8	57.9	31.7	28.7	24.8	14.8	SJL
LN CD8 <sup>+</sup>	14.0	19.3	18.3	48.4	37.1	26.7	19.0	17.2	
Thy CD4 <sup>+</sup> 8 <sup>+</sup>	16.4	17.9	17.9	47.8	30.9	35.5	19.1	14.5	
Thy CD4 <sup>+</sup> 8 <sup>-</sup>	21.9	21.9	17.2	39.0	44.2	27.4	15.8	12.6	SJL.E $\alpha$
LN CD4 <sup>+</sup>	22.9	14.6	9.4	53.1	32.5	27.7	19.3	20.5	
LN CD8 <sup>+</sup>	21.5	16.5	23.1	38.9	30.0	17.7	36.9	15.4	

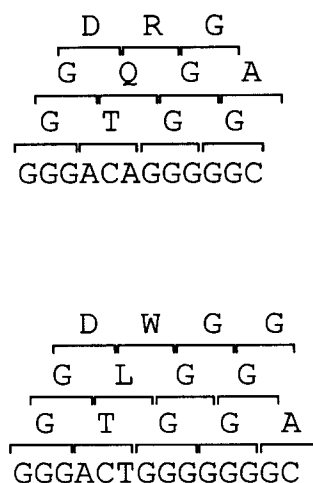
**THYMIC 4+8+**

**Figure 2.** Protein sequences. Amino acid sequences are deduced from the nucleotide sequences and displayed in the standard one-letter code. Bold letters indicate amino acids specified by the N, D, and N bases; normal characters correspond to V- or J-encoded amino acids. The nucleotide sequences from which these sequences are derived have been given accession numbers X61756–X61764 in the EMBL/GenBank Data Libraries.



**Figure 3.** Lengths of the CDR3 loops. Each histogram plots the number of sequences vs. the size of the CDR3 loops, deduced from the data sets. The definition of the CDR3 loop is based on the homology of TCR $\beta$  chains with Igs (21–23, 35). Following reference 21, CDR3 starts downstream from the conserved Ser at position 94, and ends immediately upstream from the conserved Phe at position 108 (numbering according to reference 21); thus, . . . CAS←CDR3→FGXG . . .

In Table 5, we have grouped amino acids according to the frequency at which they are encoded in the D segments of  $V_{\beta}17^{+}$  TCRs, and have weighed each one's occurrence according to the number of codons that specify it. Proline, glu-



**Figure 4.** Coding potential of the D $\beta$  segments. The translation of D $\beta$ 1 (top) and D $\beta$ 2 (bottom) segments is shown, in all reading frames.

tamic acid, tryptophan, and aspartic acid seem to be selected for; conversely, isoleucine, tyrosine, methionine, cysteine, leucine, and alanine appear to be selected against. In general, there is a preference for charged and polar residues over hydrophobics.

## Discussion

### General Properties of $V_{\beta}17^{+}$ TCRs

We have sequenced >600 transcripts that specify TCRs carrying the  $V_{\beta}17$  variable region. From this large data base, certain conclusions can be drawn about the “typical” junctional region of  $V_{\beta}17^{+}$  TCRs. Some of the conclusions are particularly interesting in light of past observations on Ig receptors.

**J $\beta$  Usage.**  $V_{\beta}17^{+}$  TCRs do not randomly use joining segments. Perhaps most striking is the finding that J $\beta$ 2s are used about four times as often as J $\beta$ 1s. This preference is evident with both expressed and nonexpressed transcripts, indicating that it is independent of repertoire selection events. One explanation may lie in the structure of the individual recombination signals: all of the heptamer and nonamer signals 5' of the J $\beta$ 1 segments are separated by 13 nucleotides (22), while those 5' of the J $\beta$ 2s are all 12 nucleotides apart (23). This difference, though small, may render the J $\beta$ 1 segments less recognizable by the recombination machinery. In fact, it has already been shown that just such a single nucleotide difference can measurably reduce the ability of an artificial substrate to undergo recombination in pre-B cells (24).

A preference for J $\beta$ 2s was foretold by previous reports that, no matter which  $V_{\beta}$  segment they carry, TCR cDNAs bearing J $\beta$ 2 fragments are isolated more frequently from recombinant libraries (25), and TCR genes with rearranged J $\beta$ 2 fragments are detected more frequently on Southern blots (26). Intriguingly, skewed J $\beta$  usage was not observed in fetal T lymphocytes (27), but we have seen it in neonatal T cells (27a).

More novel is the observation that the J $\beta$ 2s, themselves, are subject to uneven usage: J $\beta$ 2.1, J $\beta$ 2.5, and J $\beta$ 2.6 being particularly favored. Again, this nonrandom use of joining

**Table 4.** Amino Acid Frequency in NDN-encoded Stretches

G	A	S	T	L	I	V	M	F	Y	W	H	R	K	D	E	Q	N	C	P		
31.3	2.7	4.9	6.5	2.7	0	2.7	0	1.1	0	4.9	2.7	1.9	1.1	3.2	3.8	4.3	1.1	1.1	14.0	V $\beta$ 17b	B6
31.6	2.9	4.0	8.6	3.4	0	2.9	0	1.7	1.1	4.0	1.7	12.6	1.1	12.1	2.3	1.7	0.6	0.6	6.9	Thymic 4 <sup>+</sup> 8 <sup>+</sup>	
32.1	3.4	4.3	7.7	3.4	0.8	2.5	0.8	3.0	0.4	2.1	2.1	11.1	1.7	7.2	2.6	4.3	2.1	0	7.7	Thymic 4 <sup>+</sup> 8 <sup>-</sup>	SJL
32.8	3.4	5.2	9.1	2.8	2.4	1.0	0.7	2.4	1.0	1.4	2.1	7.7	2.1	8.0	2.4	4.5	1.0	0.3	9.1	Lymph node 4 <sup>+</sup>	
30.1	5.1	4.7	7.6	6.2	1.1	3.3	0	0.7	0	8.3	1.1	6.9	0.7	8.3	3.6	2.9	2.2	0	7.2	Lymph node 8 <sup>+</sup>	
32.3	2.9	3.6	8.0	2.9	1.8	2.5	0	1.5	0.3	2.9	1.5	10.2	1.8	8.7	2.5	5.8	1.5	0	9.1	Thymic 4 <sup>+</sup> 8 <sup>+</sup>	
36.8	4.0	4.9	10.7	2.7	0.4	2.2	1.3	1.3	0.4	4.4	0.8	7.5	0.8	8.8	2.7	4.0	0.8	0	4.9	Thymic 4 <sup>+</sup> 8 <sup>-</sup>	SJL.E $\alpha$
32.5	3.8	4.2	9.0	6.1	0	1.4	0	0.5	0	1.4	0.9	9.0	1.4	12.3	3.3	3.3	0.5	0	10.8	Lymph node 4 <sup>+</sup>	
28.8	4.2	4.0	5.7	5.4	0.8	2.5	0.3	2.5	0.6	5.3	0.6	8.5	0.6	9.3	2.8	1.4	0.3	0	5.9	Lymph node 8 <sup>+</sup>	

segments is true of V $\beta$ 17b TCRs, indicating that it is not dictated (entirely) by selection events. Rather, one is led to consider influences from the accessibility of chromatin, the efficiency of individual recombination signals, or the capacity of different segments to partake in secondary rearrangements.

An extensive study of human Ig heavy chain CDR3 regions

has recently been published (27). Over 100 sequences were produced from the PBL of six individuals, making no selection for particular heavy chains. The six possible J segments were also used distinctly nonrandomly in these receptors, JH4 being found in over half the sequences and JH1 and JH2 in <2%.

**N Nucleotide Addition.** On average, V $\beta$ 17<sup>+</sup>TCRs carry three N nucleotides per sequence; they may have as few as zero or as many as 14. The likelihood of finding N nucleotides is the same at the VD and DJ junctions, but the base composition at each seems to differ. There is an elevated GC content at the VD joint (with a marked enrichment in Cs), but the four bases occur more or less equally at the DJ joint.

Human Ig heavy chains show a somewhat different profile of N additions (e.g., 27; and Fig. 5 A). First of all, more Ns are added per sequence: an average of 6.2 and a maximum of 50. Second, at both the VD and DJ junctions the N additions are GC rich, but G and C are represented roughly equally. These variations can not be attributed to particularities of the V $\beta$ 17<sup>+</sup> TCR or of the human IgH data bases because a smaller set of V $\beta$ 5<sup>+</sup> TCR (28) and murine IgH (29) sequences show similar differences.

Since the distinct profile of N nucleotide additions is characteristic of V $\beta$ 17a TCRs as a whole and of unselectable V $\beta$ 17b TCRs, it is difficult for one to invoke a contribution from repertoire selection events. Rather, some feature of the recombination machinery must be responsible. It is not known with certainty what enzyme(s) is (are) responsible for the addition of N nucleotides to either TCR or Ig gene segments. Nevertheless, several facts point to terminal deoxynucleotidyl transferase (TdT).<sup>1</sup> For example, this is the only enzyme presently known to be capable of template-independent nucleotide additions (30). In addition, there seems to be a good correlation between the cell subsets that express TdT and those that support rearrangement events, including N nucleotide

**Table 5.** Amino Acid Usage in the NDN Stretch

No. of occurrences in D (Fig. 5)	No. of occurrences in stretch (all V $\beta$ 17a sequences)	Weighted no. of occurrences
0	P 176	44
	S 98	16
	E 64	32
	V 52	13
	F 40	20
	H 32	16
	K 28	14
	N 25	12
	I 21	7
	Y 10	5
	M 8	8
	C 2	1
1	R 205	34
	L 90	15
	W 88	88
	Q 79	39
2	D 193	96
	T 177	44
	A 82	20
14	G 708	118

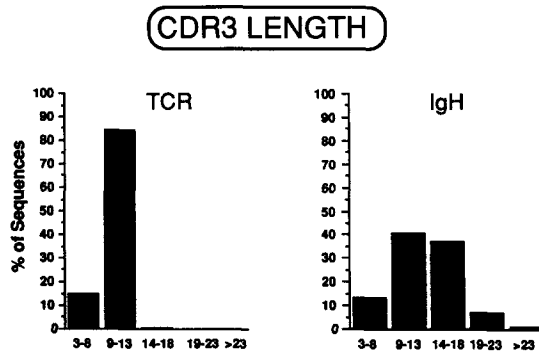
<sup>1</sup> Abbreviation used in this paper: TdT, terminal deoxynucleotidyl transferase.



A

N Nucleotide Additions					
V $\beta$ 17+ TCRs			IgH chains		
Average # of N additions	3		6.2		
Range	0-14		0-50		
Base Composition	VD	DJ	VD	DJ	
G	19.6	31.2	28.3	33.5	
A	17.0	28.1	19.8	20.2	
T	15.6	23.8	21.5	18.4	
C	48.1	16.7	30.3	27.8	
G+C	67.7	47.9	58.6	61.3	

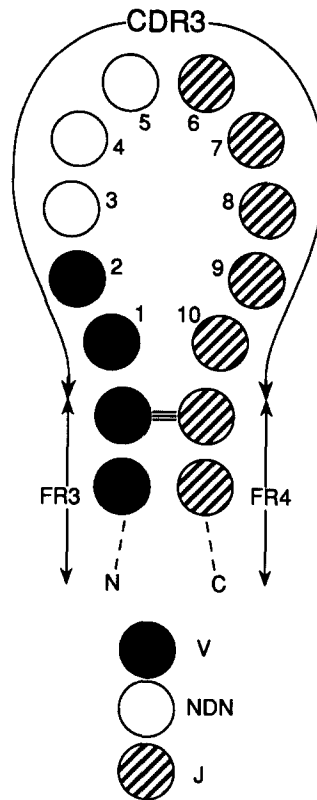
B



**Figure 5.** Comparison of TCR  $\beta$  and IgH CDR3 regions. (A) N nucleotide addition. The panel compares the average number, range, and base composition of N nucleotide additions in TCR  $\beta$  chain sequences (this paper) and IgH chains (from reference 26). (B) CDR3 length. Frequency histograms of CDR3 lengths, defined as in Fig. 4, for V $\beta$ 17+ TCR  $\beta$  chains (this paper) and IgH chains (from reference 26).

addition (31, 32). Finally, when an expressible TdT cDNA was introduced into a pre-B cell line, the frequency of N nucleotides added to a plasmid substrate increased, though not to the anticipated level (33). According to Alt and Baltimore (34), TdT prefers to polymerize dGTP. This would be consistent with the elevated GC content of N additions at the TCR VD junction and both IgH junctions. Given the contribution of Cs to this over-representation at the TCR VD junction, one is led to suggest that, in this case, N nucleotides are polymerized from the 3' end of the antisense strand of D rather than from the 5' end of the sense strand of V. Why this asymmetry is not observed at the IgH junction is mysterious and, together with the observation that the TCR DJ junction is not GC rich at all, prompts one to hypothesize that there are cell type-specific components to the recombination machinery. Certainly, transfected TCR segments can recombine in cells of the B cell lineage (32), but it is not known whether the products are, in detail, like those that are produced in T cells.

**CDR3 Lengths and Structure.** V $\beta$ 17+ TCRs have a rather narrow distribution of CDR3 lengths: average, 10; range, 6-15. As illustrated in Fig. 5 B, this distribution is much



**Figure 6.** A typical CDR3 loop in a V $\beta$ 17+ TCR $\beta$  chain. The diagram depicts a typical CDR3 loop, 10 amino acids in length. The shading indicates the most frequent (80% or greater) origin of each amino acid, and hence the degree of its variability. The dichotomy between VNDN- and J-derived faces tends to be preserved in shorter or longer CDR3 loops simply by addition or of removal amino acids on both sides. The boundaries of CDR3 and of framework regions (FR3, FR4) are defined as above. Thin horizontal lines indicate the hydrogen bonds that may hold the stem of the CDR3 loop, as hypothesized by Chothia et al. (35).

tighter than what was observed with IgH chains: long CDR3 loops (>13 amino acids), which are very common in Igs, are virtually absent in V $\beta$ 17a+ TCRs. These differences follow from their distinct profiles of N nucleotide addition, but also depend on the relative degree of exonuclease nibbling. The J segments actually contribute limited variability, because of their homology and skewed usage. The narrower histogram of TCR CDR3 lengths seems logical given that TCRs are obliged to interact with MHC molecules carrying peptide antigens, while Igs interact with antigens as diverse as small haptens and large globular proteins. Thus, one is tempted to suggest an active size selection, either during differentiation or throughout evolution. The former possibility seems unlikely, however, because V $\beta$ 17b TCRs also have a tight distribution of CDR3 lengths.

Perhaps reflecting this restrained variability, the V $\beta$ 17+ TCRs show a fairly well conserved structural feature in the CDR3s. If one plots the CDR3 amino acids onto the loop structure suggested by Chothia et al. (35), one sees that one side of the loop consists of amino acids contributed by the V and the NDN stretches, while the other is composed, in the vast majority of sequences, of amino acids encoded by the J segment (Fig. 6). Consequently, one side of the loop (and in particular, positions (3, 4, and 5) would seem to be much more variable than the other; perhaps the two sides interact with distinct components of the MHC molecule/peptide antigen duplex. Not surprisingly, this feature is not evident in the CDR3s of IgH chains.

### Selection of $V_{\beta}17$ TCRs

The sequences that we have generated derive from nine different sorted populations. Three of them can be considered unselected: the C57Bl/6 cells because  $V_{\beta}17b$  transcripts do not give rise to cell surface protein, and the thymus  $CD4^+CD8^+$  cells from  $E^-$  and  $E^+$  mice because they have low levels of TCR (and were sorted as such) and are generally considered to be an immature, pre-selection subset (36, 37). The other six populations have undergone selection of various types and allow us to view the effects of selection forces on the repertoire.

**$E^-$  and  $E^+$  Mice.** Almost all T cells carrying  $V_{\beta}17a^+$  TCRs are capable of responding to the E complex plus an unidentified B cell ligand (14). To avoid autoreactive cells,  $E^+$  mice eliminate  $V_{\beta}17^+$  T cells from their repertoire (17). But this is never complete, so that when one compares  $E^-$  with  $E^+$  mice, one sees a reduction from 8–10%  $V_{\beta}17^+$  T cells in the peripheral  $CD4^+$  compartment to 2%. A question that has interested us is whether the remaining cells have distinctive junctional region features.

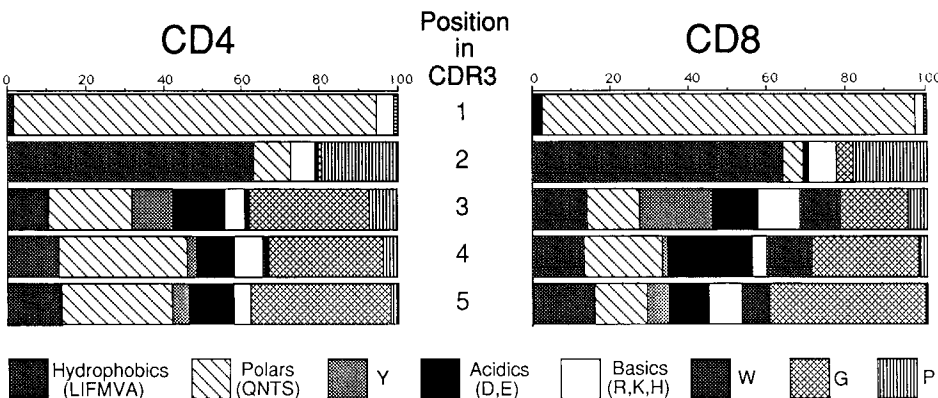
Of all the parameters analyzed, only J region usage consistently shows variability between  $E^-$  and  $E^+$  mice. Two J $\beta$ s are involved. J $\beta$ 1.1 is highly enriched in the  $V_{\beta}17^+$   $CD4$  single-positive cells of  $E^-$  mice, suggesting a positive selection event mediated by one of the SJL MHC molecules, probably A<sup>s</sup> since it is the  $CD4^+$ , not the  $CD8^+$ , cells that are involved. An increased level of J $\beta$ 1.1 is not seen in the corresponding cells of  $E^+$  mice, implying that cells expressing this J $\beta$  undergo efficient E-mediated negative selection. J $\beta$ 2.5, on the other hand, is highly enriched in the  $V_{\beta}17^+$   $CD4$  single-positive cells of  $E^+$ , but not  $E^-$ , mice. Fully one-third of the peripheral  $CD4^+$ s express this J segment. This result suggests that cells expressing J $\beta$ 2.5 are preferentially positively selected by the E molecule, are more resistant to negative selection, or both.

Such influences by MHC-mediated selection events were not really expected. The dogma has been that particular V $\beta$ s are negatively selected irrespectively of  $\alpha$  chain and junctional region contributions. This assertion is consistent with the mapping of TCR residues involved in negative selection mediated by MHC molecules in conjunction with various superantigens (16, 38, 39) but is not supported by the observation

that only partial intrathymic clonal deletion is observed in mls-positive transgenic mice that express only the  $\beta$  chain from an mls-reactive clone (40, 41, and K. Signorelli, unpublished results).

**$CD4^+$  and  $CD8^+$  Cells.** T cells expressing CD4 and CD8 are selected by different MHC molecules in the thymus (class II vs. class I) and recognize foreign antigens in the context of different MHC molecules on APC (again, class II vs. class I). Robey et al. (42) have recently reported that ectopic expression of CD4 on peripheral  $CD8^+$  lymphocytes enables at least some of them to react allogeneically across an MHC class II difference, suggesting that some TCRs selected on class I molecules have the potential to interact with class II molecules. Nonetheless, we have questioned whether junctional region sequences somehow predicate the class I/class II restriction dichotomy.

Some dissimilarities do emerge from a comparison of the  $V_{\beta}17^+$  TCRs of peripheral  $CD4^+$  and  $CD8^+$  cells. (a) Although, in general, CDR3 lengths show a relatively tight distribution, the greatest variation was observed between lymph node  $CD4^+$  and  $CD8^+$  cells (9.3 vs. 10.7 residues on average). Long CDR3 loops ( $\geq 12$ ) are significantly more frequent in CD8 than in CD4 cells (27% vs. 13% of sequences). (b) Interesting differences become apparent when we plot the frequency of individual amino acids or of amino acid groups at each position in the CDR3 loop (Fig. 7). Positions 1 and 2 are dominated by leucine and serine residues contributed by the 3' end of the  $V_{\beta}17$  segment; proline (contributed by P or N nucleotides) is also quite frequent at position 2. Positions 3, 4, and 5 (as noted above) are far more variable. The most striking difference between sequence sets is the frequent presence of tryptophan in the CDR3s of  $CD8^+$  cells and its virtual absence in those of  $CD4^+$  cells. Close to one-third of the class I-restricted TCRs carry a tryptophan at positions 3, 4, or 5 of the CDR3. This observation should be considered in light of the nature of Trp, the bulkiest of amino acids, usually found buried in hydrophobic cores. Other differences are more subtle. Polar amino acids and glycine are quite frequent in the CDR3s of  $CD4^+$  cells; charged residues are preferentially acidic at all positions. Polar amino acids and glycine are less prevalent in the CDR3s of  $CD8^+$  cells, particularly at position 3. The distribution of charged amino



**Figure 7.** Amino acid frequencies at individual positions in CDR3. Each horizontal bar depicts the frequency of amino acids or classes thereof at each of the first five positions of CDR3, within sequences derived from  $CD4^+$  or  $CD8^+$  lymph node cells. For greater representativity, we pooled data from SJL and SJL-E $\alpha$  mice (for  $CD4$ ;  $n = 113$ , for  $CD8$ ,  $n = 140$ ). The differences we observe between sequences from  $CD4^+$  or  $CD8^+$  cells at positions 3 and 4 are statistically significant ( $\chi^2$ ,  $p < 0.01$ ).

acids shows a distinct profile: while positive or negative charges are equally represented at positions 3 and 5, there is a very strong bias for negative charges at position 4.

At present, we can only speculate whether these differences between class I- and class II-restricted TCRs will prove general, and not solely restricted to  $V\beta 17^+$   $\beta$  chains. A preliminary search of the literature suggested that the differential use of Trp might indeed be general (although there were some exceptions, such as in reference 43). We do not know whether any of the differences are related to requirements for contacts with class I vs. class II molecules, or with potentially different types of peptides presented by these molecules.

### Implications

The repertoire of  $V\beta 17^+$  T cells actually used in the pe-

ripheral immune system is much smaller than the theoretical repertoire calculated by assuming entirely random operation of the recombination machinery and random export of mature cells from the thymus. J region usage is skewed, as is D region usage. The appearance of P nucleotides is nonrandom, as is the addition of N nucleotides. The overall length of the VDJ junctional region seems rather tightly controlled. We assume that these observations are not peculiar to  $V\beta 17^+$  TCRs, and actually have preliminary evidence to support this assumption from  $V\beta 6$  junctional region sequences (U. Hartwig and M. Bogue, unpublished results). Estimates of repertoire size may need to be adjusted accordingly.

We thank P. Bohn, P. Gerber, and S. Vicaire for technical assistance, and A. Staub and F. Ruffenach for oligonucleotide synthesis.

This work was supported by institutional funds from the INSERM and CNRS, and by grants to D. Mathis and C. Benoist from the National Institutes of Health and the Association pour la Recherche sur le Cancer. S. Candéas received fellowships from the Ministère de la Recherche et de la Technologie, the Association pour la Recherche sur le Cancer, and the Université Louis Pasteur.

Address correspondence to Diane Mathis, Laboratoire de Génétique Moléculaire des Eucaryotes, Unité 184 de Biologie Moléculaire et de Génie Génétique de INSERM, Institut de Chimie Biologique, 11, rue Humann, 67085 Strasbourg, France. S. Candéas' present address is the National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver, CO 80206.

Received for publication 13 May 1991 and in revised form 22 July 1991.

### References

1. Davis, M. 1988. T cell receptor genes. In *Molecular Immunology*. B. Hames, and D. Glover, editors. IRL Press, Oxford. 61-79.
2. Wilson, R., E. Lai, P. Concannon, R. Barth, and L. Hood. 1988. Structure, organization and polymorphism of murine and human T-cell receptor  $\alpha$  and  $\beta$  gene families. *Immunol. Rev.* 101:149.
3. Lafaille, J., A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa. 1989. Junctional sequences of T cell receptor  $\delta\gamma$  genes: implication for T cell lineages and for a novel inter-mediator of V(D)-J joining. *Cell*. 59:859.
4. Davis, M., and P. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature (Lond.)*. 334:395.
5. Behlke, M., H. Chou, K. Huppi, and D. Loh. 1986. Murine T cell receptor mutants with deletions of  $\beta$ -chain variable region genes. *Proc. Natl. Acad. Sci. USA*. 83:767.
6. Haqqi, T.M., S. Banerjee, G. Anderson, and C.S. David. 1989. RIII S/J (H-2<sup>i</sup>). An inbred mouse strain with a massive deletion of T cell receptor  $V\beta$  genes. *J. Exp. Med.* 169:1903.
7. Kotzin, B.L., V. Barr, and E. Palmer. 1985. A large deletion within the T-cell receptor  $\beta$ -chain gene complex in NZW mice. *Science (Wash. DC)*. 229:167.
8. Wade, T., J. Bill, P. Marrack, E. Palmer, and J.W. Kappler. 1988. Molecular basis for the nonexpression of  $V\beta 17$  in some strains of mice. *J. Immunol.* 141:2165.
9. MacDonald, H.R., A.L. Glasebrook, R. Schneider, R.K. Lees, H. Pircher, T. Pedrazzini, O. Kanagawa, J.-F. Nicolas, R.C. Howe, R.M. Zinkernagel, and H. Hengartner. 1989. T-cell reactivity and tolerance to Mls<sup>a</sup> encoded antigens. *Immunol. Rev.* 107:89.
10. Blackman, M., J. Kappler, and P. Marrack. 1990. The role of the T cell receptor in positive and negative selection of developing T cells. *Science (Wash. DC)*. 248:1335.
11. Le Meur, M., P. Gerlinger, C. Benoist, and D. Mathis. 1985. Correcting an immune response deficiency by creating E $\alpha$  gene transgenic mice. *Nature (Lond.)*. 316:38.
12. van Ewijk, W., Y. Ron, J. Monaco, J. Kappler, P. Marrack, M. Le Meur, P. Gerlinger, B. Durand, C. Benoist, and D. Mathis. 1988. Compartmentalization of MHC class II gene expression in transgenic mice. *Cell*. 53:357.
13. Benoist, C., and D. Mathis. 1989. Positive selection of the T cell repertoire: when and where does it occur? *Cell*. 58:1027.
14. Kappler, J., T. Wade, J. White, E. Kushnir, M. Blackman, J. Bill, N. Roehm, and P. Marrack. 1987a. A T cell receptor  $V\beta$  segment that imparts reactivity to a class II major histocompatibility complex product. *Cell*. 49:263.
15. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Labora-

tory, Cold Spring Harbor, NY.

16. Cazenave, P.-A., P.N. Marche, E. Jouvin-Marche, D. Voegtli, F. Bonhomme, A. Bandeira, and A. Coutinho. 1990. V $\beta$ 17 gene polymorphism in wild-derived mouse strains: two amino acid substitutions in the V $\beta$ 17 region greatly alter T cell receptor specificity. *Cell*. 63:717.
17. Kappler, J., N. Roehm, and P. Marrack. 1987. T cell tolerance by clonal elimination in the thymus. *Cell*. 49:273.
18. Zuniga-Pflucker, J.C., D.L. Longo, and A.M. Kruisbeek. 1989. Positive selection of CD4<sup>+</sup>CD8<sup>+</sup> T cells in the thymus of normal mice. *Nature (Lond.)*. 338:76.
19. Iwabuchi, K., I. Negishi, H. Arase, C. Iwabuchi, K. Ogasawara, R.A. Good, and K. Ono. Positive selection of a T-cell subpopulation in the thymus in which it develops. *Proc. Natl. Acad. Sci. USA*. 86:5089.
20. Blackman, M.A., P. Marrack, and J. Kappler. 1989. Influence of the major histocompatibility complex on positive thymic selection of V $\beta$ 17a<sup>+</sup> T cells. *Science (Wash. DC)*. 244:214.
21. Kabat, E., T. Wu, M. Reid-Miller, H. Perry, and K. Gottesman. 1987. Sequences of Immunological Interest, 4th edition. Department of Health and Human Services, Bethesda, MD.
22. Gascoigne, N., Y. Chien, D. Becker, J. Kavaler, and M. Davis. 1984. Genomic organization and sequence of T-cell receptor  $\beta$ -chain constant- and joining-region genes. *Nature (Lond.)*. 310:387.
23. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Goverman, 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  $\beta$  polypeptide. *Cell*. 37:1101.
24. Hesse, J., M. Lieber, K. Mizuuchi, and M. Gellert. 1989. V(D)J recombination: a functional definition of joining signals. *Genes & Dev*. 3:1053.
25. Toyonaga, B., and T.W. Mak. 1987. Genes of the T-cell antigen receptor in normal and malignant T cells. *Annu. Rev. Immunol.* 5:585.
26. Yamada, M., R. Wasserman, B.A. Reichard, S. Shane, A.J. Caton, and G. Rovera. 1991. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. Exp. Med.* 173:395.
27. Born, W., J. Yague, E. Palmer, J. Kappler, and P. Marrack. 1985. Rearrangement of T-cell receptor  $\beta$ -chain genes during T-cell development. *Proc. Natl. Acad. Sci. USA*. 82:2925.
- 27a. Bogue, M., S. Candéas, C. Benoist, and D. Mathis. 1991. A special repertoire of  $\alpha$ : $\beta$  T cells in neonatal mice. *EMBO (Eur. Mol. Biol. Organ.)*. In press.
28. Bill, J., O. Kanagawa, J. Linten, Y. Utsunomiya, and E. Palmer. 1990. Class I and class II MHC gene products differentially affect the fate of V $\beta$ 5 bearing thymocytes. *J. Mol. Cell. Immunol.* 4:269.
29. Feeney, A. 1990. Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. *J. Exp. Med.* 172:1377.
30. Bollum, F. 1974. Terminal deoxynucleotidyl transferase. In *The Enzymes*. Vol. X. P. Boyer, editor. Academic Press, New York. 145-171.
31. Desiderio, S., G. Yancopoulos, M. Paskind, E. Thomas, M. Boss, N. Landau, F. Alt, and D. Baltimore. 1984. Insertion of N regions into heavy chain genes correlated with expression of terminal deoxytransferase in B cells. *Nature (Lond.)*. 311:752.
32. Yancopoulos, G.D., T.K. Blackwell, H. Suh, L. Hood, and F.W. Alt. 1986. Introduced T cell receptor variable region gene segments recombine in pre-B cells: evidence that B and T cells use a common recombinase. *Cell*. 44:251.
33. Landau, N.R., D.G. Schatz, M. Rosa, and D. Baltimore. 1987. Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. *Mol. Cell. Biol.* 7:3237.
34. Alt, F., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-J $\eta$  fusions. *Proc. Natl. Acad. Sci. USA*. 79:4148.
35. Chothia, C., D.R. Boswell, and A.M. Lesk. 1988. The outline structure of the T-cell  $\alpha\beta$  receptor. *EMBO (Eur. Mol. Biol. Organ.)*. 7:3745.
36. Nikolic-Zugic, J. 1991. Phenotypic and functional stages in the intrathymic development of  $\alpha\beta$  T cells. *Immunol. Today*. 12:65.
37. Boyd, R.L., and P. Hugo. 1991. Towards an integrated view of thymopoiesis. *Immunol. Today*. 12:71.
38. Pullen, A.M., T. Wade, P. Marrack, and J. Kappler. 1990. Identification of the region of T cell receptor  $\beta$  chain that interacts with the self-superantigen Mls-1<sup>a</sup>. *Cell*. 61:1365.
39. Choi, Y., A. Herman, D. DiGiusto, T. Wade, P. Marrack, and J. Kappler. 1990. Residues of the variable region of the T-cell-receptor  $\beta$ -chain that interact with S. Aureus toxin superantigens. *Nature (Lond.)*. 346:471.
40. Blackman, M.A., H. Gerhard-Brugert, D.L. Woodland, E. Palmer, J.W. Kappler, and P. Marrack. 1990. A role for clonal inactivation in T cell tolerance to Mls-1a. *Nature (Lond.)*. 345:540.
41. Pircher, H., T.W. Mak, R. Lang, W. Ballhausen, S. Rüedi, H. Hengartner, R.M. Zinkernagel, and K. Bürki. 1989. T cell tolerance to Mls-a encoded antigens in T cell receptor V $\beta$ 8.1 chain transgenic mice. *EMBO (Eur. Mol. Biol. Organ.)*. 8:719.
42. Robey, E., F. Ramsdell, J. Elliott, D. Raulet, D. Kioussis, R. Axel, and B.J. Fowlkes. 1991. Expression of CD4 in transgenic mice alters the specificity of CD8 cells from allogeneic major histocompatibility complex. *Proc. Natl. Acad. Sci. USA*. 88:608.
43. Danska, J.S., A.M. Livingstone, V. Paragas, T. Ishihara, and C.G. Fathman. 1990. The presumptive CDR3 regions of both T cell receptor  $\alpha$  and  $\beta$  chains determine T cell specificity for myoglobin peptides. *J. Exp. Med.* 172:27.