The sizes of the CDR3 hypervariable regions of the murine T-cell receptor β chains vary as a function of the recombined germ-line segments

CHRISTOPHE PANNETIER, MADELEINE COCHET, SYLVIE DARCHE, ARMANDA CASROUGE, MARGOT ZÖLLER, AND PHILIPPE KOURILSKY

Unité de Biologie Moléculaire du Gène, Unité 277, Institut National de la Santé et de la Recherche Medicale, Institut Pasteur 25 Rue du Dr. Roux, 75724 Paris, Cédex 15, France

ABSTRACT A method using PCR amplification and primer extension with fluorescent oligonucleotides was developed to analyze T-cell repertoires. The sizes of the hypervariable CDR3-like regions of the murine T-cell antigen receptor β chains were measured for all possible V_β-J_β combinations. This analysis shows that β chains are distributed into at least 2000 groups, a value that provides a lower limit to their complexity. The CDR3 sizes appear to be dependent on the J_β and especially the V_β segment used and correlates with amino acid sequence motifs in the corresponding CDR1 region. This feature of T-cell receptors is discussed.

The specific receptors of T lymphocytes (T-cell antigen receptors; TCRs) are membrane-bound heterodimers, two types of which, $\alpha\beta$ and $\gamma\delta$, have been identified so far (1–4). The gene segments encoding the α, β, γ, and δ chains have been cloned in several species (surveyed in ref. 5). These segments are rearranged by mechanisms similar to those that operate in the immunoglobulin genes. Thus, the genes encoding the TCR α and β chains are produced by the combination of the V_α, J_α, and C_α or V_β, J_β, D_β, and C_β segments, respectively. Junctional diversity is generated by a variety of mechanisms (6), in which terminal deoxynucleotidyl-transferase is known to play a role a few days after birth (reviewed in ref. 7). The three-dimensional structure of TCRs is still unknown, but sequence homologies, the conservation of key amino acids, and modelization studies have suggested that TCRs have an antibody-like structure (2, 8–10). Accordingly, complementarity determining region (CDR) 1-, 2-, and 3-like regions have been defined in TCRs (5). The CDR3-like regions (encoded by the V-J or V-D-J junctions) are the only ones to display extensive diversity.

The diversity of TCRs, although still unknown, is thought to be as large as that of antibodies (2, 4). Repertoire analyses have been hampered by the small number of specific reagents available. Thus, monoclonal antibodies directed against several variable (V) segments have been isolated, but very few anticlonotypic antibodies have been obtained. Nucleotide sequences of rearranged genes and transcripts encoding functional chains are being accumulated in relatively large numbers, but it is difficult to envision how extensive sequencing could provide a global picture of the repertoire with the tools available to date. We have developed a method that allows a detailed, yet global, description of the repertoire of mouse TCR β-chain transcripts. We present data that describe the size spectrum of the CDR3-like region of all possible V_β-J_β combinations. Our results provide a minimal estimate of 2000 for the number of rearranged TCR β chains. They also reveal that the length of the CDR3 region correlates

with the V_β and the J_β segments. This surprising feature of murine TCRs is discussed.

MATERIALS AND METHODS

Mice. BALB/c and C57BL/6 mice were bred in the local facilities of the Pasteur Institute. C57BL/10.A, BALB.B, and BALB.K strains were from Harlan Olac (Bicester, U.K.).

Oligonucleotides and Fluorescent Dye Labeling. Oligonucleotides were synthesized by using an Applied Biosystems DNA synthesizer. Fluorescent dye labeling (with Fam, Joe, Tamra, or Rox) was performed as recommended by the supplier (Applied Biosystems). Dye-labeled oligonucleotides were purified by ethanol precipitation to eliminate unreacted dye, followed by reverse-phase chromatography to remove the unlabeled oligonucleotides. The primer sequences are given in Table 1.

RNA and cDNA Preparation. RNA was prepared with the guanidinium isothiocyanate procedure followed by CsCl ultracentrifugation (16).

Single-strand cDNA synthesis was performed by using the Boehringer Mannheim cDNA synthesis kit. Briefly, 10 μg of total RNA was incubated for 10 min at 70°C with (dT)$_{15}$ (5 μM) and each dNTP at 1 mM. After cooling, Boehringer Mannheim buffer I and avian myeloblastosis virus reverse transcriptase were added together with 40 units of RNasin (Promega). The solution was then incubated 1 hr at 43°C, ethanol-precipitated, and resuspended in 20 μl of H$_2$O.

PCR Amplification. Ten microliters of the previous synthesis, corresponding to the reverse transcription of 5 μg of total RNA, was used in the amplification. A mixture containing 50 mM KCl, 10 mM Tris·HCl (pH 8.2), 1.5 mM MgCl$_2$, 0.01% gelatin, 200 μM dNTP, 0.01% Triton X-100, the newly synthesized cDNA (or water in negative controls), Taq polymerase (Promega) (2 units/100 μl), and 0.25 μM C_β primer was distributed in 23 tubes (47 μl per tube), each containing 2.5 μl of a 5 μM V_β-specific primer solution. The solutions were overlaid with 50 μl of mineral oil. Amplifications were performed in a Perkin–Elmer or LEP thermocycler. The amplification started with a denaturation step of 1 min at 94°C, followed by 40 cycles consisting of 70 sec at 94°C, 1 min at 60°C, and 4 min at 72°C (to minimize intra-PCR recombination), and finally a 10-min step at 72°C.

Primer Extension in Run-Off Reactions. A mixture containing 50 mM KCl, 10 mM Tris·HCl (pH 8.2), 1.5 mM MgCl$_2$, 0.01% gelatin, 200 μM dNTP, 0.01% Triton X-100, Taq polymerase (Promega) (2 units/100 μl), and 25 μl of one of the previously amplified solution was distributed in 12 tubes (9 μl per tube), each containing 1 μl of a 1 μM J_β-specific labeled primer solution. The solutions were overlaid with 10 μl of

Abbreviations: CDR, complementarity determining region; MHC, major histocompatibility complex; TCR, T-cell antigen receptor.

Table 1. Oligonucleotide sequences

V gene segment-specific oligonucleotides*			J gene segment-specific oligonucleotides[†]		
V segment	Sequence (5' to 3', coding strand)	Distance to residue 95, bp	V segment	Sequence (5' to 3', noncoding strand)	Distance to residue 106, bp
$V_\beta 1$	CT GAA TGC CCA GAC AGC TCC AAG C	83	$J_\beta 1.1$	XAC TGT GAG TCT GGT TCC TTT ACC	29
$V_\beta 2$	TC ACT GAT ACG GAG CTG AGG C	74	$J_\beta 1.2$	XA AAG CCT GGT CCC TGA GCC GAA G	25
$V_\beta 3.1$	CCT TGC AGC CTA GAA ATT CAG T	63	$J_\beta 1.3$	XCT TCC TTC TCC AAA ATA GAG C	17
$V_\beta 4$	GCC TCA AGT CGC TTC CAA CCT C	102	$J_\beta 1.4$	XGA CAG CTT GGT TCC ATG ACC G	26
$V_\beta 5.1$	CAT TAT GAT AAA ATG GAG AGA GAT	135	$J_\beta 1.5$	XG AGT CCC CTC TCC AAA AAG CG	19
$V_\beta 5.2$	AAG GTG GAG AGA GAC AAA GGA TTC	126	$J_\beta 1.6$	XT CAC AGT GAG CCG GGT GCC TGC	31
$V_\beta 5.3$[‡]	AG AAA GGA AAC CTG CCT GGT T	113	$J_\beta 2.1$	XGT GAG TCG TGT TCC TGG TCC GAA G	26
$V_\beta 6$	CT CTC ACT GTG ACA TCT GCC C	56	$J_\beta 2.2$	XC CAG CAC TGT CAG CTT TGA GC	34
$V_\beta 7$	TAC AGG GTC TCA CGG AAG AAG C	90	$J_\beta 2.3$	XGT TCC TGA GCC AAA ATA CAG CG	17
$V_\beta 8.1$	CAT TAC TCA TAT GTC GCT GAC	141	$J_\beta 2.4$	XGT GCC CGC ACC AAA GTA CAA G	17
$V_\beta 8.2$	CAT TAT TCA TAT GGT GCT GGC	141	$J_\beta 2.5$	XGT GCC TGG CCC AAA GTA CTG G	17
$V_\beta 8.3$	T GCT GGC AAC CTT CGA ATA GGA	127	$J_\beta 2.7$	XC TAA AAC CGT GAG CCT GGT GC	34
$V_\beta 9$	TCT CTC TAC ATT GGC TCT GCA GGC	57			
$V_\beta 10$	ATC AAG TCT GTA GAG CCG GAG GA	48			
$V_\beta 11$	G CAC TCA ACT CTG AAG ATC CAG AGC	64			
$V_\beta 12$	G ATG GTG GGG CTT TCA AGG ATC	117			
$V_\beta 13$	AGG CCT AAA GGA ACT AAC TCC CAC	78			
$V_\beta 14$	AC GAC CAA TTC ATC CTA AGC AC	68			
$V_\beta 15$	CCC ATC AGT CAT CCC AAC TTA TCC	87			
$V_\beta 16$	C ACT CTG AAA ATC CAA CCC AC	58			
$V_\beta 17$[‡]	AG TGT TCC TCG AAC TCA CAG	80			
$V_\beta 18$	C AGC CGG CCA AAC CTA ACA TTC TC	82			
$V_\beta 19$[‡]	CT GCT AAG AAA CCA TGT ACC A	74			
$V_\beta 20$	TC TGC AGC CTG GGA ATC AGA A	62			

X, fluorescent dye. The sequence of the *C* gene segment-specific oligonucleotide (C_β) used is 5'-CTT GGG TGG AGT CAC ATT TCT C-3' (noncoding strand). The C_β sequences were derived from ref. 14.
*The V_β gene segment nomenclature and sequences were derived from ref. 3 for all segments except $V_\beta 18$ (11) and $V_\beta 20$ (12).
[†]The J_β gene segment nomenclature and sequences were derived from refs. 13–15.
[‡]Nonfunctional in BALB/c.

mineral oil. Primer extension started with a denaturation step of 2 min at 94°C, followed by 1 min at 60°C, and was stopped by an incubation of 15 min at 72°C.

Electrophoresis. Five microliters of a 95% (vol/vol) formamide/10 mM EDTA loading solution was added to 10 μl of the labeled probes. Two microliters of this solution was loaded on a 6% acrylamide/8 M urea gel. The gel was run for 5 hr in 90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3, on a 373A DNA sequencer (Applied Biosystems).

Data Analysis. Size determination of the run-off products was performed automatically using software written for this purpose and a set of size standards consisting of five run-off products derived from amplifications of five plasmids constructed as follows: cDNAs corresponding to a mixture of V_β chains were amplified with our V_β-specific primer set, and the resulting DNA was cloned at an *Eco*RV site in pBS SK(+) (Bluescript). After electroporation, several clones were isolated and sequenced. A set of five clones was retained: clone 6-K uses $V_\beta 6$, $J_\beta 1.5$; clone 6-H uses $V_\beta 6$, $J_\beta 1.6$; clone 7-R uses $V_\beta 7$, $J_\beta 2.1$; clone 7-S uses $V_\beta 7$, $J_\beta 2.2$; clone 5.8 uses $V_\beta 5.2$, $J_\beta 2.5$; and they yield DNA fragments of 96, 114, 140, 157, and 176 nt, respectively.

RESULTS

Experimental Approach. As described in detail elsewhere (ref. 17; C.P., M.C., and P.K., unpublished results), total RNA (extracted from a mixture of at least 10^5 and usually 10^6–10^7 lymphocytes) is converted into cDNA, which is then amplified by PCR using a C_β primer and one of the oligonucleotide primers specific for each of the known V_β segments in BALB/c mouse. Each amplified product is used as a template for elongation reactions initiated with each of the 12 J_β-specific oligonucleotides, which are labeled with a fluo-

rescent tag. The fluorescent run-off products, which have been elongated through CDR3 regions of various sizes, are loaded on polyacrylamide gels and subjected to electrophoresis in an automated DNA sequencer. Appropriate size standards and an analysis with our own software yield the size of the fluorescent DNA molecules. Given the position of the V_β and J_β primers (Table 1), the length of the CDR3-like region can be deduced first in nucleotides and, when in-frame, in amino acids. Following Kabat *et al.* (5) the CDR3-like region was taken as encompassing amino acids 95–106.

Estimate of the Diversity of TCR β-Chain Rearrangements. RNA from thymuses of 7-day-old BALB/c mice was analyzed as described above, using 21 V_β primers and the 12 fluorescent J_β primers. The 21 × 12 = 252 fluorescent elongation products were analyzed by electrophoresis. Representative examples are shown in Figs. 1 and 2. In all the combinations, the elongation products are separated into 6–11 discrete peaks, which are spaced by 3 nt. Comparison with size standards shows that they correspond to in-frame transcripts. Thus, each peak most likely corresponds to at least one rearranged, functional β transcript. Accordingly, a minimum of 2000 distinct rearrangements are detected in the thymus of a day 7 BALB/c mouse.

When we compared day 1 and day 7 thymocytes, we found (Fig. 1) that for all V_β-J_β combinations, the CDR3 of day 1 thymocytes was, on the average, 1 aa shorter than that of day 7 thymocytes. Nevertheless, the profiles displayed a similar number of peaks. Similarly, adult (3 months old) BALB/c thymus, spleen, and lymph nodes yielded patterns identical to those obtained with day 7 thymuses (except that all rearrangements involving $V_\beta 3$, $V_\beta 5.1$, $V_\beta 5.2$, $V_\beta 7$, $V_\beta 9$, and $V_\beta 11$, which are known to be poorly represented in the periphery of BALB/c mice, were weakly detectable in adult
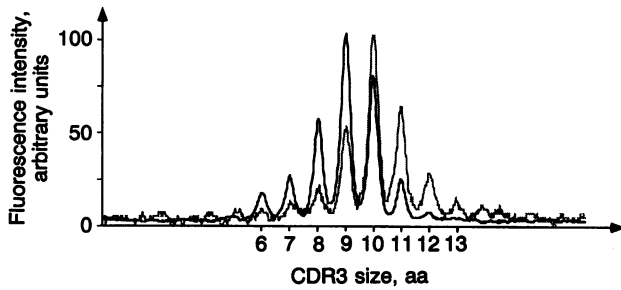
FIG. 1. The CDR3 size distribution of the TCR β chain is Gaussian and varies between 1 and 7 days of age. Total RNA was extracted from the thymus of 10 BALB/c mice at birth or 1 week later. cDNA was synthesized and, in this particular example, amplified with primers specific for the $V_\beta4$ and C_β gene segments. Aliquots of these two amplifications were subjected to run-off reactions using $J_\beta2.1$ primers labeled with two different fluorophores. The run-off products were mixed, loaded on a gel on a 373A DNA sequencer (Applied Biosystems), and run for 5 hr. Fluorescent electrophoresis profiles are displayed. For each detected peak, the corresponding CDR3 size was determined as described in *Materials and Methods*. Solid line, birth; dotted line, day 7.

peripheral organs, though giving rise to the same number of peaks) (data not shown). Finally, a comparison of the adult thymuses of BALB/c $(H\text{-}2^d)$, BALB/B $(H\text{-}2^b)$, and BALB/K $(H\text{-}2^k)$, which differ only at the level of their major histocompatibility complex (MHC), revealed very similar profiles (Fig. 2).

In summary, we find a minimum of 2000 distinct, rearranged, and in-frame β transcripts, irrespective of the age of the mouse (newborn, 7 days, or 3 months), the organ, or the MHC. For any given V_β-J_β combination, the size profiles are always similar, except that the CDR3 sizes are, on the average, 1 aa shorter in the newborn.

Clustering of V_β Segments According to the Mean Size of the CDR3 Region. We then deduced from the data the absolute sizes of the CDR3 regions. Surprisingly, we found that they are not randomly distributed. Instead, they vary as a function of the V_β-J_β combination. For example, in the $V_\beta4$-$J_\beta2.3$

combination, CDR3 regions are between 7 and 8 and 14 and 15 aa long, while in the $V_\beta3$-$J_\beta1.2$, they are between 3 and 4 and 10 and 11 residues long.

To further pursue this observation, we took advantage of the fact that the observed distributions are sufficiently regular (i.e., always Gaussian) to allow an average size to be defined. The CDR3 average sizes were found to vary between 7 and 11 aa. Using a code of five graded shades, we plotted CDR3 average sizes as a function of the V_β-J_β combination. It is apparent in Fig. 3A that each combination has a distinct average size that is related mostly to V_β usage. For example, β chains built up with $V_\beta1$ have long CDR3s, while those using $V_\beta2$ have short CDR3s.

By switching lines and columns, we found a clustering in which both V_β and J_β segment usage correlates with the CDR3 mean size (Fig. 3B). Their contributions seem to be largely independent and additive, as indicated by the presence of distinct linear domain boundaries in the matrix. The major contribution was provided by V_β usage: for a given J_β segment, the V_β usage could change the mean size by up to 4 aa, whereas for a given V_β segment, the J_β usage contribution could modify the mean size by 1 or 2 aa residues at the most.

DISCUSSION

We have established an approach that allows for an overall analysis of the T-cell repertoire. It was used here to analyze the repertoire of TCR β chains in BALB/c mice. We demonstrate, first of all, that the repertoire of TCR β-chain transcripts is composed of a minimum of 2000 distinct groups characterized by the usage of a given V_β segment, a given J_β segment, and a given size for the CDR3-like region. Taking into account that for any given V_β-J_β pairs the individual size peaks can be heterogeneous in sequence (ref. 17; unpublished results), the β-chain repertoire probably includes many more than 2000 members. Given the large number of J_α segments and the known V_α-J_α junctional variability, it appears likely that a similar lower limit would be obtained for α chains. Thus, as a conservative estimate based on simple probability theory, the actual $\alpha\beta$ repertoire should contain millions of specificities.
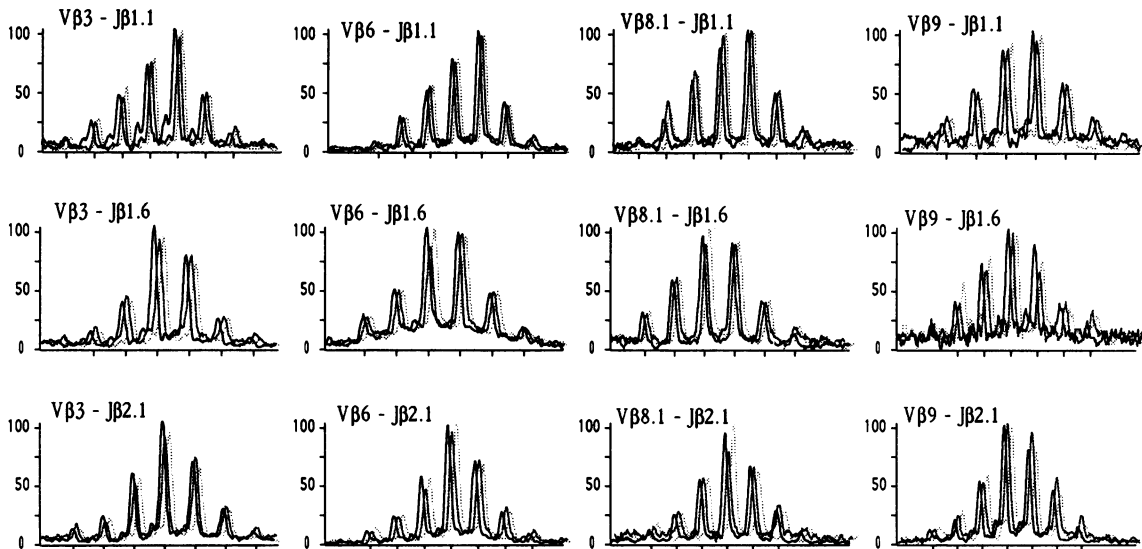

FIG. 2. TCR β-chain CDR3 mean sizes in adult BALB/B, BALB/c, and BALB/K thymuses. Total RNA was extracted from the thymuses of adult BALB/B, BALB/c, and BALB/K congenic mice displaying different MHC haplotypes. cDNA was synthesized and amplified in 6 parallel experiments with a C_β-specific primer and primers specific for six V_β gene segments. For each amplification, 12 run-off experiments were then performed in parallel, using the 12 different J_β-specific fluorescent primers. The run-off products were analyzed on a 373A Applied Biosystems sequencer. The fluorescent profiles are displayed as in Fig. 1 (x axis, CDR3 size; y axis, fluorescence intensity) for 12 of the analyzed combinations. To provide a clearer view, the BALB.K profile (dotted line) was displaced to the right of the BALB/c profile (dashed line), and the BALB.B profile (solid line) was placed slightly to the left of the BALB/c profile.
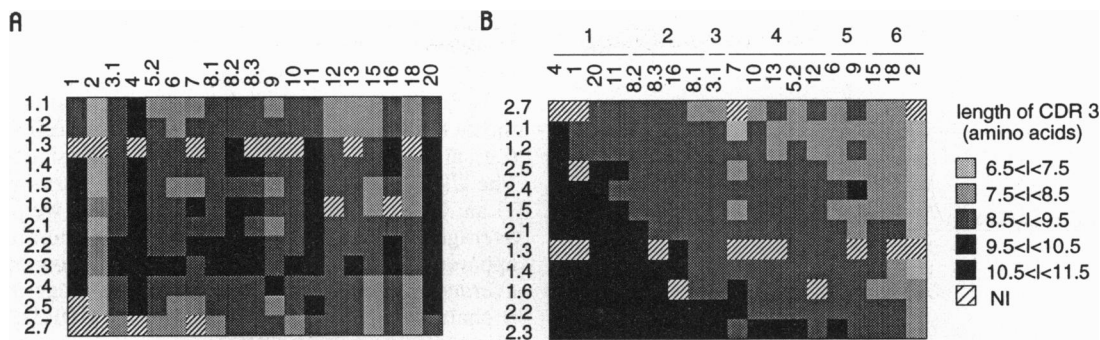
FIG. 3. TCR β-chain CDR3 mean size in 7-day-old BALB/c thymus varies with V_β and J_β usage. (*A*) Total RNA was extracted from the thymuses of ten 7-day-old BALB/c mice. cDNA was synthesized and amplified in 19 parallel experiments with a C_β-specific primer and a different primer specific for each of the functional V_β gene segments ($V_\beta 5.1$ and *-14* were not tested). Twelve run-off experiments using each of the 12 J_β-specific fluorescent primers were then carried out in parallel on each of these PCR products. The run-off products were analyzed in a 373A Applied Biosystems sequencer. CDR3 mean sizes were calculated as described in *Materials and Methods*. (*B*) The V_β and J_β gene segments were ordered so as to cluster the different CDR3 mean sizes into connected domains. NI (not interpretable) refers to weak fluorescence signals that were reproducibly obtained in a few V_β-J_β combinations, such that the size distributions could not be safely analyzed above the background. We assume that the β-chain transcripts displaying these combinations are poorly represented. The numbers at left indicate the J_β gene segments, and those along the top indicate the V_β gene segments. The numbers 1–6 above the V_β gene segment numbers in *B* refer to the CDR1 group numbers (see Table 2).

The size of the CDR3-like regions of the β chains could not previously be studied systematically for every single V_β-J_β combination. Our analysis of β transcripts shows that most of them are actually derived from in-frame genetic rearrangements. We find that, in any V_β-J_β combination, CDR3 sizes can vary by as many as 7–10 aa and that in healthy (nonimmunized) animals, the size distributions are invariably very regular [i.e., Gaussian or quasi-Gaussian (Figs. 1 and 2)]. This feature has allowed us to define an average CDR3 size for each observed distribution. The underlying assumption that the distribution of the fluorescent run-off products reflects that of the starting mRNA population is, in fact, strongly supported by a set of separate experiments, which demonstrates that the series of enzymatic reactions yielding the fluorescent elongation products is highly quantitative (18, 19). It is therefore convenient to plot only the CDR3 average sizes (rather than the size distributions) for any given V_β-J_β combination, as was done with a graded code in Fig. 3*A*.

The results in Fig. 3*A* establish that the CDR3 average size varies from 7 to 11 aa as a function of the V_β-J_β combination. This relationship is the same in central and peripheral lymphoid organs of adult BALB/c mice. It does not change in young *versus* adult or old animals, and it is not dependent upon MHC polymorphism, since BALB/c (*H-2^d*), BALB/B (*H-2^b*), and BALB/K (*H-2^k*) mice display similar profiles (Fig. 2). Furthermore, spleens of adult C3H (*H-2^k*), C57BL/6 (*H-2^b*), and C57B/10A (*H-2^a*) mice show only a few differences (apart from those related to the total or partial I-E or Mls dependent deletions) (data not shown). However, the CDR3 mean sizes vary during ontogeny, since thymocytes isolated from newborn (Fig. 1) and 17-day-old fetuses (data not shown) display TCR β-chain CDR3 regions that are, on the average, 1 aa shorter than those in day 7 or adult thymocytes. This confirms and extends earlier findings (7, 20). As noted by Bogue *et al.* (20) and Feeney (7), the increase in CDR3 sizes between day 1 and day 7 might correlate with the activation of terminal deoxynucleotidyltransferase in thymocytes. The shift takes place at a time when tolerance to self is learned by the animal. Its functional significance remains mysterious and deserves further investigation.

It is interesting to note that the observed CDR3 lengths are significantly shorter than those estimated from germinal segments. The calculations could be made only for the seven genomic V_β sequences available to us. In the V_β-$D_\beta 1$-J_β combination of germ-line segments, values of about 12 aa are obtained for $V_\beta 3.1$, $V_\beta 5.2$, $V_\beta 8.1$, $V_\beta 8.2$, $V_\beta 8.3$, and $V_\beta 11$

and 11 aa for $V_\beta 18$. These figures are 3 and 4 aa larger than the CDR3 mean sizes observed in adults and the newborn, respectively. In other terms, the germ-line configurations would fall at the very upper end of the experimental distributions and do not account for a major fraction of the observed β chains. It may be asked whether TCR displaying such germ-line configurations are selected against during ontogeny, and if so, which selection pressure operates on the genomic germ-line segments.

The patchwork displayed in Fig. 3*A* was changed into a more regular pattern when the V_β and, to a lesser extent, the J_β segments were ordered according to the CDR3 mean sizes, as demonstrated in Fig. 3*B*. Similar to the relationship between CDR3 mean sizes and V_β-J_β usage, this clustering does not vary much between mouse strains. It is also conserved during ontogeny, since CDR3 mean sizes are shifted by about 1 aa in all combinations.

We have attempted to find a basis for the relationship between CDR3 mean sizes and the use of V_β and J_β segments that might account for this clustering. The CDR3 mean sizes do not correlate in any obvious way with the genomic organization of the V_β and J_β families. By comparing the nucleotide sequences of the V_β and J_β sets, we found no correlation with the size patterns. In particular, the five V_β subgroups defined by Chothia *et al.* (9) do not correlate with CDR3 size distributions. When we considered the amino acid sequences, we observed that the V_β segments could be organized into six groups sharing sequence homologies within their CDR1-like regions (Table 2). Group 1 displays a sequence close to XLGHNA; the alanine residue at position 31 is characteristic of that group. Group 2 displays TNNHNY. Group 3 contains only one member ($V_\beta 3.1$), which has proline at position 30. In group 4, ISGHDT is found; alanine at position 31 is excluded. Group 5 displays XXNHDT, and group 6 includes the three V_β segments in which the CDR1 length is 7 aa (rather than 6 aa) and in which glutamine is found at position 29. We found no other correlations in the framework or in other variable regions (Table 2). The J_β segments that result in a leucine at position 106 tended to correlate with the longer CDR3s (Table 2).

Although it is impossible to assess the significance of these correlations without further structural and/or functional information, the observation as such is striking and deserves discussion. Three major hypotheses can be considered.

(*i*) The bell-shape distribution of sizes observed in every situation for each of the V_β-J_β combinations suggests that

Table 2. Correlation between CDR3 average length and V_β usage

CDR1 subgroup	V segment*	CDR1 sequence,[†] aa 26–31	CDR2 sequence,[†] aa 48–63	HVR 4 sequence,[‡] aa 70–74	Most homologous V segment[§]	Homology with the following V segment[§]
1	$V_\beta4$	YLGHNA	SYSYQKLMDNQTASS	SKKNH	$V_\beta1$ (43)	$V_\beta1$ (43)
	$V_\beta1$	HLGHNA	LYNLKQLIRNETUPS	PDSSK	$V_\beta4$ (43)	$V_\beta20$ (32)
	$V_\beta20$	EKGHTA	YFNNQQPLDQIDMUK	PSSSL	$V_\beta3$ (57)	$V_\beta11$ (43)
	$V_\beta11$	ISGHSA	YFRNQAPIDDSGMPK	PNQSH	$V_\beta12$ (56)	$V_\beta8.2$ (27)
2	$V_\beta8.2$	INHHHN	YSYGAGSTEKGDIPD	PSQEH	$V_\beta8.1$ (77)	$V_\beta8.3$ (68)
	$V_\beta8.3$	INSHNY	YSYGAGNLQIGDUPD	TTQED	$V_\beta8.1$ (64)	$V_\beta16$ (21)
	$V_\beta16$¶	USNHLY	NFYNUKUMEKSKLFK	PDGSY	$V_\beta12$ (40)	$V_\beta8.1$ (26)
	$V_\beta8.1$	INNHDY	YSYUADSTEKGDIPD	PSQEH	$V_\beta8.2$ (77)	$V_\beta3.1$ (23)
3	$V_\beta3.1$	EKGHPU	NFQHQEULQQIDMTEK	PSNSP	$V_\beta17$ (58)	$V_\beta7$ (17)
4	$V_\beta7$	DMSHET	ISYDUDSNSEGDIPK	KKREH	$V_\beta8.1$ (42)	$V_\beta10$ (21)
	$V_\beta10$	TLGHDT	SYNNKQLIUNETUP	SDKAH	$V_\beta1$ (43)	$V_\beta13$ (32)
	$V_\beta13$	ISGHDT	YFRDEAUIDNSQLPS	PGKTN	$V_\beta11$ (45)	$V_\beta5.2$ (29)
	$V_\beta5.2$	ISGHSN	QHYEKUERDKGFLPS	FDDYH	$V_\beta5.1$ (74)	$V_\beta12$ (32)
	$V_\beta12$	USGHND	YFRSKSLMEDGGAFK	LNNSF	$V_\beta11$ (45)	$V_\beta6$ (25)
5	$V_\beta6$	NFNHDT	YSITENDLQKGDLSE	EKKSS	$V_\beta8.1$ (38)	$V_\beta9$ (25)
	$V_\beta9$	TNNHDT	FYYDKILNREADTF	PNNSF	$V_\beta8.1$ (30)	$V_\beta15$ (25)
6	$V_\beta15$	UGFQATS	STUNSAIKYEQNFTQE	PNLSF	$V_\beta18$ (36)	$V_\beta18$ (36)
	$V_\beta18$	ADSQUUS	ANEGSEATYESGFTKD	PNLTF	$V_\beta15$ (36)	$V_\beta2$ (21)
	$V_\beta2$	KNSQYPU	TLRSPGDKEUKSLPGA	UTDTE	$V_\beta13$ (24)	

*For V_β gene segments nomenclature, see Table 1.

[†]CDR1 and CDR2 boundaries were defined according to ref. 9, following Kabat's nomenclature (5).

[‡]Hypervariable region 4 (HV4) boundaries were taken from ref. 22.

[§]The murine V_β gene segments were aligned at the amino acid level using Corpet's software (23). Numbers in brackets represent the pairwise alignment score for the two given V_β segments.

¶Residue 52 is valine in C57BL mice and glycine in SJL mice (5).

there are increasing deviations (in nucleotides or amino acids) from a statistically "optimal" CDR3 length. This might be a property of the recombination machinery. If such is the case, the machinery should also be endowed with the ability to recognize which V_β-J_β pair is being used. We have not been able to identify nucleotide sequences that would serve as signals for this recognition. These could have escaped our attention, but the apparent correlation between CDR3 mean sizes and CDR1 amino acid sequences has yet to be explained.

(*ii*) The CDR3 mean sizes might be selected for the TCR to assume an appropriate structure. For example, differences in the CDR1 sequences might influence the β-chain structure, either directly or indirectly (e.g., via its association with the α chain), in a way that would require the CDR3 region to acquire the right dimension.

(*iii*) Finally, the CDR3 mean sizes might be selected in such a way that functional MHC–peptide–TCR complexes could be formed. As an example (21), one could imagine that the V_β CDR1 sequence might specifically recognize particular non-polymorphic regions of MHC molecules (given our observation that size patterns are not influenced by the MHC haplotype). These "anchor" points might, to some extent, dictate the geometry of the MHC–peptide–TCR complex for which a minimum of six topologies, corresponding to our six V_β subgroups, might exist. Depending upon the anchor points and the orientation of the β chain, longer or shorter CDR3 might be required to contact the area with the most frequent epitopes.

1. Kronenberg, M., Siu, G., Hood, L. & Shastri, N. (1986) *Annu. Rev. Immunol.* **4**, 529–591.
2. Davis, M. M. & Bjorkman, P. J. (1988) *Nature (London)* **334**, 395–402.
3. Wilson, R. K., Lai, E., Concannon, P., Barth, R. K. & Hood, L. E. (1988) *Immunol. Rev.* **101**, 149–172.
4. Ashwell, J. D. & Klausner, R. D. (1990) *Annu. Rev. Immunol.* **8**, 139–167.
5. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C., eds. (1991) *Sequences of Proteins of Immunological Interest* (Natl. Inst. Health, Bethesda, MD).
6. Alt, F. W., Oltz, E. M., Young, F., Gorman, J., Taccioli, G. & Chen, J. (1992) *Immunol. Today* **13**, 306–314.
7. Feeney, A. J. (1991) *J. Exp. Med.* **14**, 115–124.
8. Novotny, J., Tonegawa, S., Saito, H., Kranz, D. M. & Eisen, H. N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 742–746.
9. Chothia, C., Boswell, D. R. & Lesk, A. M. (1988) *EMBO J.* **7**, 3745–3755.
10. Claverie, J.-M., Prochnicka-Chalufour, A. & Bougueleret, L. (1989) *Immunol. Today* **10**, 10–14.
11. Louie, M. C., Nelson, C. A. & Loh, D. Y. (1989) *J. Exp. Med.* **170**, 1987–1998.
12. Six, A., Jouvin-Marche, E., Loh, D. Y., Cazenave, P.-A. & Marche, P. N. (1991) *J. Exp. Med.* **174**, 1263–1266.
13. Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. (1984) *Nature (London)* **309**, 322–326.
14. Gascoigne, N. R. J., Chien, Y.-H., Becker, D. M., Kavaler, J. & Davis, M. M. (1984) *Nature (London)* **310**, 387–391.
15. Malissen, M., Minard, K., Mjolsness, S., Kronenberg, M., Goverman, J., Hunkapiller, T., Prystowsky, M. B., Yoshikai, Y., Fitch, F., Mak, T. W. & Hood, L. (1984) *Cell* **37**, 1101–1110.
16. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
17. Cochet, M., Pannetier, C., Regnault, A., Darche, S., Leclerc, C. & Kourilsky, P. (1992) *Eur. J. Immunol.* **22**, 2639–2647.
18. Pannetier, C., Cochet, M., Darche, S. & Kourilsky, P. (1992) *C.R. Acad. Sci. Ser. 3* **315**, 271–277.
19. Pannetier, C., Delassus, S., Darche, S., Saucier, C. & Kourilsky, P. (1993) *Nucleic Acids Res.* **21**, 577–583.
20. Bogue, M., Candeias, S., Benoist, C. & Mathis, D. (1991) *EMBO J.* **10**, 3647–3654.
21. Kourilsky, P. & Claverie, J.-M. (1989) *Adv. Immunol.* **45**, 107–193.
22. Jores, R., Alzari, P. M. & Meo, T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9138–9142.
23. Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881–10890.