SEQUENCES AND DIVERSITY OF HUMAN T CELL RECEPTOR $\beta$ CHAIN VARIABLE REGION GENES

By NOBUHIRO KIMURA,* BARRY TOYONAGA,* YASUNOBU YOSHIKAI,* FREDERIC TRIEBEL,† PATRICE DEBRE,‡ MARK D. MINDEN,* AND TAK W. MAK*

From *The Ontario Cancer Institute and the Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M4X 1K9; and the ‡Departement d'Hematologie, Centre Hospitalier Universitaire Pitie-Salpetriere, Paris, France

During the last two years, genes of the T cell antigen receptor (TcR) have been cloned (1-6). The $\beta$ chain, which was identified first (1, 2), is located on human chromosome 7 and murine chromosome 6 (7, 8). The germline DNA organization of both human (9) and murine (10, 11) TcR $\beta$ genes has been determined. As has been shown (12-15) in the study of Ig genes, it appears that functional TcR $\beta$ chains are encoded by genes that have undergone somatic rearrangement of noncontiguous variable (V$\beta$) diversity (D$\beta$), joining (J$\beta$), and constant (C$\beta$) gene segments. The data accumulated so far is consistent with a recombinative model for the generation of TcR $\beta$ chain diversity. Such diversity is a fundamental characteristic of the TcR $\beta$ component, since the receptor complex is thought (16) to be involved in the recognition of foreign antigen in the context of the product. An extensive recognition repertoire is made possible through combinatorial joining, junctional flexibility, and N region diversification of the germline $\beta$ gene segments. Apparently, the analogy to the generation of Ig diversity does not include somatic hypermutation (17).

The extent of germline TcR V$\beta$ gene segment multiplicity must be determined to more fully understand the operation of the TcR. The upper limit of the V$\beta$ gene segment repertoire in the mouse has been estimated (18, 19) at $\sim$20 different segments. Certain mutant mouse strains have been found (20) to carry even fewer germline V$\beta$ gene segments. The human germline V$\beta$ multiplicity is estimated in this study by analyzing 22 DNA sequences, and using Southern blots of germline DNA with some of the cDNAs as probes. The upper limit of human V$\beta$ gene segments is estimated at $\sim$100 different segments, considerably greater than in mice. A similar analysis of the human TcR $\alpha$ chain repertoire is presented elsewhere (21).

This work was supported by grants from the Medical Research Council of Canada, the National Cancer Institute of Canada, and the Natural Sciences and Engineering Research Council of Canada, and by a special research fund from the University of Toronto. B. Toyonaga and Y. Yoshikai are recipients of awards from the MRC of Canada.

1 Abbreviations used in this paper: ATL, adult T cell leukemia; ds, double-stranded; TcR, T cell receptor for antigen.
Materials and Methods

Construction of cDNA Libraries. Double-stranded (ds) cDNA was synthesized from poly(A)⁺ RNA derived from PHA-stimulated peripheral human T cells (21) or a T cell clone specific for diphtheria toxoid. After treatment with Eco RI methylase and size selection, the ds cDNA was cloned into the Eco RI site of λgt10 using Eco RT linkers as described before (22).

Isolation of Human β Chain cDNA Clones. The peripheral human T cell library was plated on E. coli C600/HFL. Screening of duplicate filters was carried out according to a standard procedure. Hybridizations were done for 18 h at 65°C in 5X SSC, 5X Denhardt’s, 100 μg/ml denatured salmon sperm DNA, and 0.5 μg ³²P-labelled nick-translated constant region fragment of JUR-β2 (a β cDNA probe that we have described previously [23]). Filters were washed in 2X SSC, 0.1% SDS several times at room temperature, followed by washing in 0.2X SSC at 65°C.

DNA Sequencing. The cDNA inserts were subcloned into the M13mp9 bacteriophage vector, and their sequences were determined using both the universal M13 primer and the specific-primer-directed dideoxynucleotide sequencing technique (24, 25).

Southern Blot Analysis. DNA was extracted from bone marrow cells and digested with Eco RI and Hind III. DNA (10 μg) was electrophoresed through 0.8% agarose and transferred to nitrocellulose filters as described by Southern (26). Hybridization was for 24 h at 65°C in 5X SSC, 5X Denhardt’s, 100 μg/ml denatured salmon sperm DNA, 10% dextran sulfate, and 0.5 μg ³²P-labelled nick-translated cDNA probe. Filters were washed at 65°C with 3X SSC/0.1% SDS.

Results

Sequence of β cDNA Clones. The repertoire of the human TcR β chain genes was examined by analysis of 22 Vβ sequences from various sources. 13 cDNAs were obtained from the screening of a PHA-stimulated human peripheral blood T cell library using a Cβ2 probe (HBP series). Another two β cDNAs were obtained from human T cell clones specific for Diptheria toxoid (DT110, DT259). The remaining seven Vβ sequences from human T cell leukemia and T cell tumor cell line sources were taken from the existing literature (see Fig. 1). Note that the four adult T cell leukemia (ATL) sequences (MT 11, ATL 121, ATL 122, and ATL 21) shown are artificially derived from the published genomic sequences by data splicing.

Examination of the 22 Vβ sequences reveals that they vary in their 5’ halves, which correspond to the variable regions of the TcR β chain genes. The 3’ cDNA sequences correspond to one of the two constant regions, Cβ1 or Cβ2. The junction points between the Vβ, Dβ, Jβ, and Cβ gene segments in the sequences were determined by comparison to previously reported human germline data (9, 13). The 19 clones found to contain Vβ gene segments are compared in Fig. 1. The corresponding deduced protein sequences are compared in Fig. 2. Spaces have been introduced into the sequences to maximize intersequence similarities. The spaces at the nucleotide level are consistent with those in the protein sequences, and vice versa. This alignment process was extended to a similar set of Vβ cDNA sequences provided to us by P. Concannon, L. Pickering, and L. Hood (personal communication) for comparative purposes. In this way, a unified nomenclature was established (i.e., the first member of family 8 is called Vβ8.1). Pairwise comparisons of all Vβ DNA sequences, as shown in Fig. 1, were performed to establish Vβ gene segment families. For the comparisons, the 3’
**Figure 1.** Nucleotide sequences of 19 human TCR β chain variable region gene sequences obtained from human T lymphocytes. Clones designated HBP were isolated from a PHA-stimulated human peripheral blood T cell cDNA library. DT110 and DT295 were obtained from human T cell clones specific for diphtheria toxoid. The remaining clones from human tumor and leukemia cell lines were taken from the literature: YT35 (1), and the identical sequence is reported for a cDNA clone isolated from cell line JM (33); HPB2 (23) (from HPB-ALL); 4D1 (31) (from HPB-MLT); MT11, ATL121, ATL122, ATL21 (17) (artificially spliced genomic sequences from cell lines of patients with ATL). Corresponding $\delta_2$ and $\delta_4$ sequences for these clones are shown in Fig. 2. Spaces have been added into the sequences to maximize similarities. Those with $\geq 75\%$ homology have been grouped into families. Deducing amino acid from the 80\% DNA consensus sequence. Variable gene segment 3' boundary based upon comparison with corresponding germine sequences are indicated by a slash (/). Cytosine residue resulting in a nonfunctional message due to a translational reading frameshift in clone 4D1 is indicated by a small c. DT295 has a deletion within $\delta_4$ 2.1 (+). See discussion regarding HBP54 and HBP34 family designation (+).
Figure 2. Deduced protein sequence of human TcR β chain variable regions from cDNA sequences (see Fig. 1 legend for details). Consensus sequences at 30, 50, 75, and 95% levels are displayed above the main figure. Nucleotides leading to nonfunctional messages due to translational reading frameshifts (4D1, ATL121) are printed in small letters and underlined.
Figure 3. VDJC region nucleotide sequences of the human TcR β chains. Vβ and Jβ boundaries, based upon comparison with germline sequences are indicated by a slash (/). (a) Sequences using Cα1 are grouped according to Jα2 use. Sequences identical to either germline Dα1.1 gene segments are shown in smaller bold capitals. DT259 has a deletion within Jα2.1 ('). Deduced amino acids from the 85% nucleotide consensus sequence, excluding genomic intron sequences of HBP15, HBP31, and HBP22. Incomplete messages resulting from partial rearrangements (9-mer/7-mer recognition sequences are underlined).
boundary of the $V_\beta$ region was taken to be two codons after the consensus codons Tyr-X-Cys-Ala. Members of a family have at least 75% homology over their overlapping sequences. Examples of similar $V_\beta$ gene segments associated with different $D_\beta$ and $J_\beta$ gene segments were found (HPB04, 4D1, and HPB50). 17 of the 19 $V_\alpha$ sequences are unique, indicating there are at least 17 different germline gene segments.

Nucleotide sequences of the region between the $V_\alpha$ and $C_\beta$ gene segments are shown in Fig. 3, a and b. The 21 $J_\beta$-containing sequences have been divided into 9 using $C_\beta1$ and 12 using $C_\beta2$. The location and identity of each $J_\beta$ gene segment within the sequences were determined very easily by comparison with previously determined germline $J_\beta$ data (9). Only 3 of the 13 known $J_\beta$ gene segments were not found in this study ($J_\beta1.4$, $J_\beta1.6$, and $J_\beta2.6$). The exact identification of the cDNA $D_\beta$ gene segments, even with the germline counterparts (see Fig. 3, a and b, bottom), was impossible due to the combined effects of junctional and N-region diversification with the limited number of bases making up the germline segment. An attempt to assign the sequences to either $D_\beta1.1$ or $D_\beta2.1$ has been made in Fig. 3, a and b, respectively.

Only 2 of the 21 rearranged sequences (DT259 and HBP48) show a recombinational event that has mixed gene segments between the two distinct $D_\beta J_\beta C_\beta$.
germline clusters. In both cases, the Dβ1.1 gene segment has rearranged to a Jβ2 gene segment. In addition, it is possible that the cDNA HBP48 arose from a recombination between both Dβ1.1 and Dβ2.1. In all of the clones, the Jβ gene segments have recombined with the Cβ gene segment located immediately downstream to it in the germline state. The cDNAs HBP15, HBP31, and HBP22 do not contain any Vβ-like sequences. HBP15 and HBP31 have resulted from an incomplete rearrangement that brought only Dβ2.1, Jβ2.1, and Cβ2 together. The nucleotide sequence 5' to the Dβ2.1 residues correspond exactly to the germline sequence associated with Dβ2.1. The putative nonamer/heptamer recombination signal sequences are underlined in Fig. 2b. HBP22 arose by transcription of a nonrearranged gene, with subsequent splicing leaving only Jβ2.4 joined to Cβ2. Again the nucleotides 5' to Jβ2.4 correspond exactly to those found in the germline 5' to Jβ2.4.

Southern Analysis of Vβ Gene Segments in Human Germline DNA. As an alternative approach in the determination of germline Vβ multiplicity, Southern blot analyses of Bam HI or Eco RI-digested germline DNA was performed using selected cDNAs as probes. Representative results are illustrated in Fig. 4. Bands associated with Cβ1 and Cβ2 are marked with arrows based on comparison with published data (9, 12, 13). In most cases, multiple Vβ-associated bands can be observed. Only HBP55 (Vβ3 family) produces a Southern blot result, which indicates that it might be the only member of the family. Familial assignments of the cDNAs based on these Southern blot results are consistent with those made by sequence analysis, as described earlier. For example, compare HBP25 and HBP50 in Figs. 1 and 4. Also, the Southern blot pattern of HBP41 (Vβ8.2) matches that of YT35 (Vβ8.1), described previously (12, 13). Based on Fig. 4, six families, and a total of at least 30 crosshybridizing members can be identified.

Discussion

22 human nucleotide sequences encoding the variable region of the TcR β chain have been analyzed in this study (Figs. 1 and 2). Three appear to be the result of incomplete rearrangement, or of transcription of unrearranged genes, since only germline intron sequences can be found where Vβ gene segments are expected (Fig. 2b). Although several sequences are truncated at the 5' end and do not contain leader and start codon sequences, only two, 4D1 and ATL121, lead to translational reading frameshift mutations. Note that 4D1 and ATL121 were isolated from nonfunctional leukemic T cell clones. Excluding these, all of the other sequences have the potential to code for functional TcR β chains.

Both of the known Dβ gene segments are used with roughly equal frequencies in the sequences examined. In two cases (DT259 and HBP48) a Dβ1.1 gene segment is found joined to Jβ2 and Cβ2 gene segments. These are the only examples of recombination of gene segments between the two known Dβ1.1Cβ germline gene segment clusters.

21 of 22 sequences contain a Jβ gene segment. Each Jβ is joined to the Cβ of the same Dβ1.1Cβ germline cluster. 10 of the 13 known Jβ gene segments occur, with no obvious preferences. In only one case (DT259) did the cDNA Jβ sequence not correspond exactly with previously reported germline data. However, the
protein sequence Phe-Gly-X-Gly, characteristic of Ig and TcR J gene segments, is conserved throughout.

Figs. 2 and 3 clearly show that junctional flexibility and N-region diversity are used fully in the recombination of Vβ, Dβ, and Jβ gene segments. No diversity is apparent at the Jβ-Cβ boundary, in contrast to the, albeit limited, Jα-Cα (21) and Jγ-Cγ (27) diversity observed.

Of the 19 Vβ-like sequences reported, 17 are unique at the nucleotide level. Two pairs differ by only one amino acid (Vβ4.2/4.3, Vβ8.1/8.3), excluding leader sequences. If they are due to polymorphism, then 15 out of 19 Vβ gene segments are unique. Assuming random expression of Vβ gene segments, an estimate of the germline Vβ gene segment repertoire can be obtained statistically (18). The data is consistent with a maximum Vβ gene segment repertoire of 104 at the 95% confidence level, and a most probable repertoire of 38. In mice, Barth et al. (18) have estimated the Vβ repertoire to have an upper bound at 21 (at 95% confidence level), and a most probable size of 13. Similarly, Behlke et al. (19) have predicted a 95% upper bound of 30 and most probable value of 18 for the mouse. This estimate is supported by Southern blots of germline DNA probed with selected cDNAs (Fig. 4). The majority of blots show several bands that crosshybridize with the Vβ probe. This is in marked contrast to a similar study with mice, in which most of the Vβ gene segments show little to no crosshybridization (19).

The human germline TcR β chain gene segment multiplicity is considerably higher than that of the human Ig λ light chain or the murine TcR β chain (16, 17, 23). It is lower than that predicted for the heavy (28) and κ (29) Ig V gene segments. The magnitude of the human Vβ gene segment repertoire appears to at least rival its human and murine α chain counterparts.

Pairwise comparisons of the Vβ sequences as aligned in Fig. 1 allowed percent homologies to be calculated over their overlapping regions. In general, the Vβ sequences homologies ranged from 30 to 100%, with most falling between 30 and 60%. The deduced protein sequences shared similar relationships, and the consensus sequences (Fig. 2) are reminiscent of other Ig-like V regions. A Kabat-Wu variability analysis (30) of the protein sequences did not define easily distinguishable regions of hypervariability, which are characteristic of Ig V regions. To a large extent, variation in Vβ chain lengths are responsible for the lack of region definition, since spaces must be introduced to maintain the structural similarities.

Even so, certain sequence similarities allowed subsets or families of Vβ segments to be defined. By definition, family members share ≥75% similarity at the DNA level. Using this criterion, the 19 Vβ sequences were divided into 11 families. The nomenclature was chosen to be consistent with those of Concannon et al. HBP54 and HBP34 Vβ gene segments are assigned to families Vβ12 and Vβ13, respectively, by comparison with data of Concannon et al. However, HBP54 and HBP34 share ≥75% similarity, and can be considered members of the same family. Thus the definition of a family is not exact, and depends on the sample size examined. The Southern blot data (Fig. 4) are consistent with the family assignments.
Based on the data of this study and Concannon et al., there are at least 16 human TcR Vβ families, each often containing several members. The total Vβ repertoire, based on number and size of families, is consistent with the estimate arrived at by statistical analysis of the sequence data described above. Again, the human TcR Vβ repertoire appears to be similar to that of human and mouse TcR Vα (this is partly due to a high number of Jα segments in both man and mouse [21, 32]) and not as small as that estimated for the mouse TcR Vβ chain.

It is possible that the number of murine TcR Vβ segments is only a reflection of the strains of mice examined. This hypothesis is supported by the findings that other strains of mice, SJL, C57/L, and C57/br are known to have fewer Vβ gene segments (20). The TcR Vβ repertoires of other murine strains, as well as wild mice, may be larger. Alternatively, the difference between the estimates may also be due to the sources of TcR cDNA sequences. If only those cDNAs obtained from a single heterogeneous (i.e., thymus, spleen, etc.) library are considered, much lower Tα redundancy is found in both human and murine cases. Differences between human (this study) and murine (19) Vβ family sizes estimated by Southern blot data could result from differences in filter washing conditions. A final possibility is that the TcR Vα gene segments of the mouse have diverged more rapidly than either Ig or TcR α chain variable gene segments.

In summary, it appears that the human TcR Vβ germline gene segment multiplicity could be three to four times greater than that reported for the mouse. We discuss some possible explanations for this discrepancy, which are consistent with the data. The estimates of Vβ gene segment family size and number are similar to that of human and mouse TcR Vα chains. An upper limit of ~100 germline Vβ gene segments is consistent with the data presented in this study. Even though the Vβ regions encoded by the cDNA sequences examined herein contain many Ig-like characteristics, they do not display the same uniformity in length or distribution of well-defined regions of hypervariability. Ig-like tertiary or quaternary structures for the TcR are not predicted. Thus the structural features necessary for the recognition of antigen in the context of the MHC product by the TcR remains a mystery.

Summary

The nucleotide sequences of 22 human T cell antigen receptor (TcR) β chain variable region genes isolated from various T lymphocytes have been analyzed. Of the 19 variable gene segment (Vβ)-containing sequences, 17 were unique. The Vβ gene segments were grouped into 11 families. Comparisons were made with the data of Concannon et al. to unify the nomenclature. The data is consistent with a total Vβ gene segment repertoire with a most probable value of 38 members and an upper bound of 104 members at the 95% confidence level. Southern blot data of germline DNA using selected TcR Vβ cDNAs as probes support this estimate. The human repertoire is approximately three to four times greater than that reported for the mouse. Explanations for this discrepancy are proposed.
We thank Maurizio Laudisa for assistance in computer analysis and Nicolette Caccia for editorial assistance.

Received for publication 28 March 1986 and in revised form 2 June 1986.

References


