Development of TCRB CDR3 length repertoire of human T lymphocytes

Junko Nishio¹, Mihoko Suzuki¹, Toshihiro Nanki¹, Nobuyuki Miyasaka¹ and Hitoshi Kohsaka¹

¹Department of Bioregulatory Medicine and Rheumatology, Graduate School, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan

Keywords: gene rearrangement, peripheral blood, thymus

Abstract

The third complementarity-determining region (CDR3) of TCR interacts directly with antigenic peptides bound to grooves of MHC molecules. Thus, it is the most critical TCR structure in launching acquired immunity and in determining fates of developing thymocytes. Since length is one of the components defining the CDR3 heterogeneity, the CDR3 length repertoires have been studied in various T cell subsets from humans in physiological and pathological conditions. However, how the CDR3 length repertoire develops has been addressed only by a few reports, including one showing that CDR3 of CD4 thymocytes becomes shorter during thymic development. Here, we explored multiple regulations on the development of the TCRB CDR3 length repertoires in the thymus and the peripheral blood. CDR3 length spectratyping was employed to examine thymocyte and peripheral T cell populations for their CDR3 length repertoires. We have found that repertoire distribution patterns depend on use of the BV gene. The BV-dependent patterns were shaped during thymic selections and maintained in the peripheral blood. Differences in the mean CDR3 length among different BV subsets were seen throughout lymphocyte development. We also observed that CDR3 was shortened in both CD4 and CD8 thymocytes. Of note, the degrees of the shortening depended on the CD4/CD8 lineage and on use of the BV gene. When expansions of peripheral T cell clones are negligible, no obvious difference was seen between mature thymocytes and peripheral lymphocytes. Thus, the TCRB CDR3 length repertoires are finely tuned in the thymus before the lymphocytes emigrate into the peripheral blood.

Introduction

Using surface receptors for antigens, αβ T cells recognize antigenic peptides bound to MHC class I or II molecules. Studies using X-ray crystallography have demonstrated that three-dimensional structures composed by the first, second and third complementarity-determining regions (CDR1, 2 and 3) of TCR α and β chains interact directly with peptides presented by the MHC molecules (1,2). Avidity of the interaction is defined by topological structure and location of charged amino acid residues of the interface peptides (3). In the TCR chain, CDR3 nucleic acid sequence is most diverse because it is generated by recombination of multiple V, D (in the case of TCRβ) and J gene segments, and by random addition of interlocating N region nucleotides (4,5). Since this region interacts most closely with the antigenic peptide, the diversity of the CDR3 amino acid sequences accounts for a wide array of antigen specificities within the functional T cell repertoire.

The molecular interaction of interface peptides is similarly important in association between antigenic peptides and MHC molecules. This interaction limits heterogeneity of peptides that can bind to the products of a given MHC allele (6). The length of the antigenic peptides is also restricted by interaction with MHC and with TCR (6). In contrast, the TCR CDR3 segments are more diverse in length. This might be explained by weaker association of antigenic peptides with TCR than with MHC (3,7). However, it remains to be seen how the CDR3 length repertoire is regulated during thymic development and in peripheral blood.

The αβ T cell repertoire develops through a number of selection steps in the thymus. TCRB gene rearrangement becomes complete first at the stage of CD3+CD4+CD8- immature single-positive (CD4 ISP) thymocytes (8,9). If their TCRB genes rearrange in-frame and their products pair successfully with pre-βα chains, these cells survive and
proliferate to become CD4+CD8+ double-positive thymocytes (8,10). They express TCRβ chains together with products of the in-frame rearranged TCRA gene. The double-positive thymocytes then undergo positive and negative selection, which make mature CD4 and CD8 T cell repertoires desirable to eliminate foreign pathogens. Although these processes are directed by the avidity of TCR with its ligand (11), their effects on the CDR3 length repertoire have hardly been explored.

The heterogeneity of the TCR CDR3 length in T cells at any developmental stages can be tested with TCR CDR3 length spectratyping (CLS). This method visualizes the distribution of TCR CDR3 length as histograms (12). It has been shown that typical histograms that are derived from mature peripheral T lymphocyte pools display a Gaussian-like distribution with 3-base spacing. If a histogram is biased by an unexpectedly high frequency at a specific length, it indicates that the studied population contains an expanded T cell clone whose CDR3 has the corresponding length. Based on this, the TCR CLS technique has been employed to study clonal perturbation of T cell repertoires from healthy donors and patients with various inflammatory diseases (13–19). The results have given us some insights into the physiology and pathology of T cell homeostasis.

The above facts all indicate the importance of discerning how heterogeneity of the CDR3 length repertoire is physiologically regulated, especially in the thymus. No gross difference in CDR3 length distribution between fetal and adult T cell pools has been reported (20). Yassai et al. (21) reported that thymocytes with shorter TCRB CDR3 are selected during transition from CD4+CD8+ thymocytes to CD4 SP thymocytes. Their subsequent report used murine systems to show that the shortening is mediated by TCR–peptide–MHC interaction in the thymus (22). Of note, they suggested that human repertoires might be under distinct regulation. Other investigators have described that different TCRB CDR3 lengths were preferred by different BV and BJ combinations in mice (23), and BJ genes in humans (24). However, no studies have addressed which stages in lymphocyte development are responsible for these differences.

How are the CDR3 length repertoires of various T cell subsets formed, modulated and maintained in the thymus? How does the shortening occur in the human thymus? The present study was conducted to address these issues. By examining thymocytes and peripheral T cells for TCR CLS patterns, we have found that formation of human TCR CDR3 length repertoires is under multiplex regulations in the thymus.

**Methods**

**Samples**

Thymic fragments and peripheral blood were collected from child donors during heart surgery for correction of congenital cardiac anomalies. They were from 1 to 13 years old (mean 5.6 years old). They suffered from no immunological or hematological disorders. Consent forms were obtained before the operation. CD4 ISP thymocytes, mature CD4 and CD8 SP thymocytes, and peripheral CD4 and CD8 T lymphocytes were sorted from the thymic tissues or peripheral lymphocytes as described previously (25). Purities of the separated cells were >94%.

**PCR**

RNA were extracted from the sorted thymocytes and lymphocytes, and converted to cDNA (25). To amplify TCR transcripts with individual TCRBV family genes, the cDNA were subjected to PCR using a fluorescent TCRBC-specific anti-sense primer (Cβb) and a panel of sense oligonucleotide primers specific to TCRBV gene families (26). The amplification reaction consisted of 35 cycles of 1 min at 94°C, 1 min at 60°C and 1 min at 72°C, with final extension at 72°C for 7 min.

To amplify TCR transcripts with individual members of the BV7 family (BV7S1, BV7S2 and BV7S3 genes), a sense primer specific to the three BV7 family genes (Vγ7os: GGA GCT CAT
GTT TGT CTA CA) and a BC-specific antisense primer [C₃b (26)] were used for primary PCR. The reaction consisted of 25 cycles of 1 min at 94°C, 1 min at 53°C and 1 min at 72°C followed by final extension at 72°C for 7 min. Part of the products were further amplified with a nested sense primer specific to BV7S1, BV7S2 or BV7S3 genes (V₃b 7S1s: TAC AGC TAT GAG AAA CTC TC; V₃b 7S2s: TAC AGT CTT GAA GAA CGG GT; or V₃b 7S3s: TCT ACA ACT TTA AAG AAC AGA C) and the fluorescent C₃b primer. The reaction consisted of 25 cycles of 1 min at 94°C, 1 min at 53°C and 1 min at 72°C followed by final extension at 72°C for 7 min.

**TCR CLS**

The PCR products were fractionated on denaturing 7% polyacrylamide gel in a Hitachi SQ-5500 sequencer (Hitachi Electronics Engineering, Tokyo, Japan). The data were analyzed with the associated software to display histograms. Relative percentage of the TCRB transcripts of a given length to total TCRB transcripts in the BV subsets, which is called the frequency in this report, was calculated by dividing the fluorescence intensity of the corresponding peaks by the sum of the intensity of all peaks.

**Statistical analyses**

CDR3 length, defined as previously described (20), ranged from 6 to 60 bases. Nineteen frequency values within this range were treated as variables for cluster analyses, which were performed with Statistica 4.1J (Tulsa, OK). The variances were calculated as follows: \( \sum_{n=2}^{19} F_{3n} \times (L_{3n} - \text{mean CDR3 length})^2 \), where \( F_{3n} \) stands for the frequency value that corresponds to a given CDR3 length of \( L_{3n} \). The Kruskal–Wallis test was used to compare the central peak (CP) frequencies, variances and mean CDR3 lengths of the histograms of different BV subsets. The Mann–Whitney test was used to compare these parameters of the histograms of the CD4 ISP thymocytes with those of the other populations.

**Results**

**BV-dependent TCRB CDR3 length repertoires of peripheral CD4 T cells**

In order to characterize unbiased TCRB CDR3 length repertoires of the mature T lymphocytes, peripheral CD4 T lymphocytes from six child donors were examined. This population was studied because biases of the T cell repertoires by clonally expanded T cells are more frequent in elderly individuals and in the CD8 T cell pool (13,15,27,28). Although the histogram of each BV subset displayed a Gaussian-like distribution without outstanding biases, different BV subsets had slightly different patterns. Histograms of BV1, 3 and 7 gene families of three donors are shown to represent such differences (Fig. 1A). The shapes of different BV subsets were distinguished by the height of the CP that always had the highest frequency and by the width of the span. The histograms of BV1 had a high CP and narrow span, those of BV3 had a low CP and wide span, and BV7 had modestly high CP and a narrow span.

The characteristics were quantitatively assessed with the CP frequencies and the variances; the variances indicate span of the histograms. These two values were calculated for all BV subsets studied (Fig. 1B). Various combinations of CP frequencies and variances were observed. Reflecting the histogram pattern of the BV3 subset, its CP frequencies were low and the variances were remarkably large. This was also the case with the BV11 subset. The BV1 subset, as well as the BV6S1 subset, had high CP frequencies and small variances. The two parameters also describe the characteristics of the BV7 subset: moderate CP frequency and small variance.

Although some BV subsets had higher CP frequencies than BV1, or smaller variances than BV1 and 7, the BV1, 3 and 7 subsets were further studied to investigate how these differences develop during T lymphocyte development. The other BV subsets occasionally had minor and random biases, which should be due to small expansions of T cell clones. The characteristics of the three BV subsets and similarity within the same subsets could be illuminated by line graphs of the CDR3 length repertoires from six donors (Fig. 2A). Statistical comparison of the CP frequencies and the variances among the three subsets from six donors demonstrated that the CP frequencies of the BV1 subset were highest, while those of the BV3 subset were lowest, and that the variances in the BV3 subset were largest (Fig. 3A and B).

Pannetier et al. (23) reported that the mean TCRB CDR3 length of murine lymphocytes depends on use of BV genes. This was the case with human peripheral lymphocytes; the mean length of the TCRBV7 transcripts was longest, while that of the TCRBV3 transcripts was shortest (Fig. 3C).

Overall differences in the CLS patterns were elucidated by cluster analysis, which treated 19 frequency values at 6–60 bases as variables. A total of 18 histograms from six donors were segregated into three groups, each of which contained histograms of BV1, 3 or 7 subsets (Fig. 4A).

According to the published database, the BV7 family consists of BV7S1, BV7S2 and BV7S3 genes, while BV1 and 3 families have a single gene member (29). The histograms of the BV7 subset were derived from the PCR products that were generated with a primer specific to all BV7 family genes. In order to examine the TCR transcripts with individual BV7 genes, these transcripts were independently amplified with specific primers. The CLS distributions of the transcripts with the three BV7 genes were homologous and no statistical differences in CP frequency, variance or mean TCR length were observed (data not shown). Thus, the BV7 family subset was analyzed as a whole in the present studies.

**Development of the BV-dependent repertoires in the thymus**

TCRBV1, 3 and 7 transcripts that were derived from CD4 ISP thymocytes, and CD4 and CD8 SP thymocytes from the same set of donors were analyzed to study how the BV-dependent characteristics develop. As was discussed in our previous report (25), the CD4 ISP thymocytes have undergone TCRB gene rearrangement, but have not started positive or negative selection. Thus, unlike CD4 CD8 double-positive cells, a part of which are already under pressure of thymic selection, they are the best for investigation of primordial TCR repertoires.

The histograms of the CD4 and CD8 SP thymocytes shared the same characteristics as those of the peripheral CD4 cells.
In both SP populations, the CP frequencies of the BV1 subset were highest and those of the BV3 subset were lowest. The variances in the BV3 subset were largest. These differences were statistically significant (Fig. 3A and B).

Peripheral CD8 T cells from the same donors were analyzed in the same way. Their histograms were often biased since CD8 T cells are prone to large clonal expansions. Nevertheless, the BV-dependent characteristics were held well by the peripheral CD8 T cells (Figs 2B, and 3A and B).

In contrast, differences among the three subsets were not significant in the histograms of the CD4 ISP thymocytes (Fig. 2B). These histograms shared the same features, which were characterized by low CP and wide span regardless of BV gene use. In all of the three BV subsets, the CP frequencies and variances of the CD4 ISP thymocytes were different from those of the SP thymocytes and peripheral T cells in a statistically significant manner (Fig. 3A and B). The CP frequencies of CD4 ISP were significantly lower than those of CD4 SP in the BV1, 3, and 7 subsets \((P < 0.01, P < 0.05\) and \(P < 0.05\) respectively), than those of CD8 SP \((P < 0.01\) for each subset), than those of peripheral CD4 \((P < 0.01, P < 0.05\) and \(P < 0.01\) respectively) and than those of peripheral CD8 \((P < 0.01, P < 0.05\) and \(P < 0.01\) respectively). The variances of CD4 ISP were significantly larger than those of CD4 SP in the

Fig. 2. TCRB CLS histograms of the BV1, 3 and 7 subsets of the peripheral lymphocytes and thymocytes. Distributions of the frequencies are presented in a line graph format. The data of the six donors (D1–D6) are shown as overlaid line graphs in each panel to illuminate BV-specific characteristics. (A) The histograms of the three BV subsets in their peripheral CD4 T cells. (B) The histograms of the CD4 ISP, CD4 SP and CD8 SP thymocytes, and peripheral CD8 T cells. Except for the CD4 ISP thymocytes, the histograms of each BV subset were similar. The BV1 and 3 subsets of the peripheral CD8 T cells from D4 were considerably biased, probably because of clonal expansions.
seen in the CD4 ISP thymocytes (Fig. 3C). These results imply that positive and negative selections exert distinct effects on CLS distribution pattern and on CDR3 length.

Cluster analyses segregated the histograms of the CD4 and CD8 SP thymocytes into three groups, each of which contained primarily those of the same BV subset (Fig. 4C and D). The histograms of the peripheral CD8 T cells also fell into the three groups except for the histograms with biases (Fig. 4B). Notably, the same analysis of the histograms of the ISP thymocytes failed to discriminate BV gene use (Fig. 4E). This should be due to similarity of the distribution patterns and suggests that the difference in length alone is not enough for segregation.

**BV- and co-receptor-dependent shortening of TCR CDR3 length in the human thymus**

It has been reported that TCR CDR3 shortens during transition from the ISP thymocytes to the SP thymocytes (21). This was observed in our studies of the mean CDR3 length; the CD4 ISP thymocytes had longer CDR3 than the other populations (Fig. 3C). To investigate this further, we analyzed plots of differences in frequency (ΔF) and skew values (ΔΣF), both of which have been defined by Yassai et al. (21,22). ΔF can be calculated by subtracting the CLS frequency of a given population from that of the other at the same length. A cluster of positive ΔF values on the right of an inflection point with a corresponding cluster of negative ΔF values on the other side indicates that the given population has shorter CDR3. ΔΣF is the sum of ΔF values to the right of the inflection points. The ΔF plots and ΔΣF were calculated by subtraction of the frequencies of the CD4 and CD8 SP thymocytes from those of the CD4 ISP thymocytes in the three BV subsets (Fig. 5A). Their patterns and positive ΔΣF values showed that both SP thymocyte populations had shorter CDR3 than the CD4 ISP thymocytes irrespective of BV subset.

Interaction of TCR with endogenous antigens dictates CD4/CD8 lineage commitment during positive and negative selections in the thymus. This led us to assume that the shortening could be a function of the lineage if it is a consequence of TCR triggering. We then calculated ΔF between cells with different lineages; between the CD4 and CD8 SP thymocytes and between the peripheral CD4 and CD8 T cells (Fig. 5B). The ΔF plots and ΔΣF showed that the TCR CDR3 length of CD4 lineage cells was shorter in the BV3 subset, whereas it was longer in the BV7 subset. No significant differences in CDR3 length were seen in the BV1 subset. Thus, differential shortening between CD4 and CD8 lineage cells was observed and it depended on BV gene use.

Additionally, the CD4 SP thymocytes and the peripheral CD4 T lymphocytes, as well as the CD8 SP thymocytes and the peripheral CD8 T lymphocytes, were compared. The results showed that CDR3 of CD4 or CD8 lineage cells do not shorten in the peripheral blood (data not shown).

**Discussion**

The present study has elucidated how TCR CDR3 length repertoires of CD4 and CD8 T cells in different BV subsets develop in the human thymus and peripheral blood. The CDR3 length repertoires had BV-dependent distribution patterns.
They were shaped during thymic selections and maintained in the peripheral blood. In contrast, the BV-dependent differences in the TCR CDR3 length were observed throughout lymphocyte development. The CDR3 became shorter during thymic selections, but did not change the BV-dependent differences seen before the selections. Finally, the degrees of the shortening differed between CD4 and CD8 lineage cells, and also were dependent on BV gene use. The repertoires of peripheral lymphocytes reflected directly those of mature SP thymocytes except for biases induced by clonal expansions.

Although it was known that different CDR3 lengths were preferred by different BV subsets, the BV-dependent distribution patterns are disclosed here for the first time. Unlike the difference in length, the different patterns become evident during positive and negative selections, accompanied by an increase of the CP frequency and narrowing of the distribution span. This argues that they are shaped under the pressure of positive and negative selections in the thymus. Most studies that employed the TCR CLS technique disregarded the differences, probably because the technique was used for identification of gross changes.

We have found that the distribution patterns of three gene members of the BV7 family shared the same characteristics. This ensures that the CLS histograms generated with the primer specific to BV7 family genes were not artifacts. In this regard, we have found that different gene members of the BV6 family could have similar distribution patterns (Fig. 1B). Also, both BV3 and 11 subsets shared histograms with low CP frequencies and large variances. Arden et al. (29) pointed out that these two genes are closely related both structurally and in their CDR3 sequences. According to their TCRBV gene classification, BV1, 3 and 7 fall into different groups. These facts argue that the BV-dependent differences could be attributable to the structure of TCRβ chains.

Using murine thymus, Pannetier et al. (23) observed that different BV subsets prefer different TCR CDR3 lengths. We found that the BV-dependent difference in mean TCR length already occurred in the CD4 ISP thymocytes. This implies that the difference is regulated by TCR rearrangement. Also, since the CD4 ISP thymocytes with complete TCRB gene rearrangement are under pressure of subsequent β selection for association with pre-Ta chains, the β selection could contribute to the difference formation. Moreover, the CLS histograms...
of the BV1, 3 and 7 subsets of the CD4 ISP thymocytes were similar, but not necessarily identical (Fig. 2B), suggesting that the rearrangement and/or β selection may have a subsidiary effect in shaping the CLS distribution patterns.

To address further if the rearrangement per se regulates the BV-dependent difference in CDR3 length, we tried to amplify non-productively recombined TCRBV1, 3 and 7 genes from peripheral T lymphocytes that do not express TCRβ1, 3 or 7. However, even from >10⁷, a sufficient amount of the rearranged genes could not be amplified for the TCR CLS analyses. This was consistent with the fact that the TCRB CLS patterns of the CD4 ISP thymocytes always had 3-base pair spacing, indicating that all transcripts were in-frame. The 3-base spacing was also observed by Yassai et al. (21) who examined TCRB genomic DNA derived from the same population. It is known that 85% of CD4 ISP thymocytes retain the TCRB germline configuration, while only 5% express rearranged TCRB gene products (9). Thus, thymocytes with out-of-frame TCRB rearrangements must be diluted out quickly by those expressing complete TCRβ/pre-Trβ and become undetectable with conventional technologies.

In separate experiments, we have assessed the mean TCR CDR3 lengths of the other BV subsets and found that the BV subsets with similar CLS patterns do not necessarily have similar CDR3 length (data not shown). The differences in shortening between CD4 and CD8 lineage cells were not a function of the distribution patterns (data not shown). Thus, distribution pattern and length appeared to be regulated independently.

Yassai et al. (21,22) reported TCR shortening in the human and murine thymi. By examining murine thymocytes for the BV1–BJ2 recombinants, they have shown that the shortening...
Development of human TCRB CDR3 length repertoire
occurs to a larger extent in the CD4 lineage cells than in the CD8 lineage cells. They failed to see differential shortening in humans and suggested a distinct regulation for human thymocytes. However, we observed a clear difference between the CD4 and CD8 T cells. We found that the differential shortening was a function of BV gene use. These data imply that the shortening in humans is regulated by antigen recognition by TCR.

The TCR shortening could be affected by allelic variations of MHC gene products, and differential shortening between CD4 and CD8 lineage cells could be due to differential orthogonal geometry of TCR and antigenic peptide in the grooves of MHC class I and II molecules (22). However, the differences in mean CDR3 length were preserved throughout the thymic selections. The CD4 and CD8 lineage cells share the same CLS distribution patterns. Thus, although limitations in sample collection did not allow us to investigate the effects of HLA variations, the geometry should not be the only factor regulating CDR3 length repertoire.

Development of the TCRB CDR3 length repertoire is regulated delicately in the thymus. Peripheral selections have little effect unless T cell clones expand massively in response to immunological insults. Elucidation of these thymic biases may shed more light on molecular interaction of TCR with self-peptide–MHC in the thymus.

Acknowledgements
This study was supported by a grant from the Ministry of Health, Labor and Welfare, Japan. The authors thank Dr Takeshi Hiramatsu at Tokyo Women’s Medical University for providing us with thymus and peripheral blood samples.

Abbreviations

ΔF differences in frequency
ΣΔF skew value
CDR complementarity-determining region
CLS CDR3 length spectratyping
CP central peak
ISP immature single positive
SP single positive

References
26 Kohsaka, H., Taniguchi, A., Chen, P. P., Ollier, W. E. R. and...


