A Mechanism for TCR Sharing between T Cell Subsets and Individuals Revealed by Pyrosequencing

Vanessa Venturi, Máire F. Quigley, Hui Yee Greenaway, Pauline C. Ng, Zachary S. Ende, Tina McIntosh, Tedi E. Asher, Jorge R. Almeida, Samuel Levy, David A. Price, Miles P. Davenport and Daniel C. Douek

*J Immunol* 2011; 186:4285-4294; Prepublished online 7 March 2011;
doi: 10.4049/jimmunol.1003898
http://www.jimmunol.org/content/186/7/4285
A Mechanism for TCR Sharing between T Cell Subsets and Individuals Revealed by Pyrosequencing

Vanessa Venturi,*1 MAire F. Quigley,†,1 Hui Yee Greenaway,*1 Pauline C. Ng,*‡ §
Zachary S. Ende, † Tina McIntosh,§ Ted E. Asher, † Jorge R. Almeida, † Samuel Levy,§
David A. Price, †,§ Miles P. Davenport, # and Daniel C. Douek†

The human naive T cell repertoire is the repository of a vast array of TCRs. However, the factors that shape their hierarchical distribution and relationship with the memory repertoire remain poorly understood. In this study, we used polychromatic flow cytometry to isolate highly pure memory and naive CD8+ T cells, stringently defined with multiple phenotypic markers, and used deep sequencing to characterize corresponding portions of their respective TCR repertoires from four individuals. The extent of interindividual TCR sharing and the overlap between the memory and naive compartments within individuals were determined by TCR clonotype frequencies, such that higher-frequency clonotypes were more commonly shared between compartments and individuals. TCR clonotype frequencies were, in turn, predicted by the efficiency of their production during V(D)J recombination. Thus, convergent recombination shapes the TCR repertoire of the memory and naive T cell pools, as well as their interrelationship within and between individuals. The Journal of Immunology, 2011, 186: 4285–4294.

Although the importance of T cells in the control of infectious agents throughout the lifetime of an individual is now well established, the influence of variation within the repertoires of Ag-specific TCRs has only more recently become appreciated (1–3). Furthermore, the relative impact of the different events involved in T cell development on shaping the peripheral T cell repertoire remains poorly understood. The processes of germline V, J, and D (for β-chains) gene recombination and junctional diversification by nucleotide deletion and addition generate a vast array of >1018 unique TCRββ sequences in humans; in turn, these nascent TCRs undergo thymic selection and peripheral expansion to populate the naive T cell pool, which is estimated to contain ~2.5 × 107 unique TCRs (4–6). Yet, because our current understanding of T cell repertoire composition derives predominantly from studies of rather restricted epitope-specific responses, it remains unclear whether the individual clonotypes (T cell populations defined by their expressed TCRs) that make up such highly diverse naive T cell pools are equally represented to produce a featureless “repertoire landscape” or whether the distribution of clonotype sizes is quantitatively varied. If the latter is the case, the question arises as to what are the mechanisms that underlie the hierarchical features of clonotype distribution and whether any such patterns are preserved and transmitted to the memory T cell pools.

The relationship between the memory and naive T cell repertoires has generally been examined within the fixed framework of epitope-specific T cells that survive an acute response to antigenic challenge and enter the memory pool. Yet, this approach should be considered in light of the dynamic relationship between these two populations, which involves continuous thymic output of new naive T cells, homeostatic maintenance of the peripheral repertoire, and recruitment of naive T cells to the memory pool through episodic and persistent antigenic stimulation (7, 8). Recruitment from the naive T cell pool seems to be highly efficient (9), such that even CD8+ T cell clonotypes with very low avidity for cognate Ag (10, 11) are mobilized from the naive repertoire. Therefore, one might expect that the clonotypic landscape of the memory T cell repertoire after initial recruitment should reflect the hierarchical distribution of available clonotypes within the naive T cell repertoire. Observations in mice have provided evidence for (12) and against (13) a role for thymic selection in shaping clonotypic output to generate a hierarchical distribution within the naive T cell pool, whereas other studies found only limited overlap between the repertoires of central and effector memory T cells in mice (14) and humans (15). However, small sample sizes and insufficient sequencing depth in some of these studies may have led to an underestimation of clonotypic frequencies and the extent of overlap.

Studies in genetically identical mice showed that almost a third of the naive peripheral TCRβ repertoire overlaps between individuals
The extent to which total peripheral TCR repertoires overlap between individuals in outbred human populations is not known. However, the interindividual sharing of epitope-specific clonotypes that has been observed in a variety of immune responses in humans (as well as in mice and rhesus macaques) (17, 18) suggests that memory and naive human TCR repertoires are also not unique to an individual. Our previous studies of such epitope-specific CD8+ T cell repertoires suggested that a process of convergent recombination enables some TCR sequences to be produced more frequently by V(D)J gene recombination than others (19–21). Convergent recombination predicts that such frequently made clonotypes will have a greater likelihood of being shared between individuals within epitope-specific memory CD8+ T cell repertoires and that they will be present more frequently in the naive T cell repertoires within and among individuals (18).

The T cell repertoire can now be probed with sufficient depth using new sequencing technology (22–24) and, when combined with high-definition polychromatic flow cytometric sorting (25), one may interrogate the clonotypic composition of stringently defined memory and naive T cell populations. In this study, we addressed the following hypotheses: that TCRδ clonotype frequency in the memory and naive CD8+ T cell pools is highly influenced by production frequency as predicted by convergent recombination; that the hierarchy of TCRβ clonotype frequencies established by V(D)J gene recombination is maintained through pairing with TCRα-chains, thymic selection, and peripheral expansion into the naive pool and subsequently into the memory T cell repertoire; and that convergent recombination is a fundamental determinant of TCRβ clonotype sharing between individuals and between the memory and naive T cell pools within individuals.

Materials and Methods

Subjects

Donors were healthy HLA-A*0201+ blood bank volunteers, who provided consent in accordance with the Institution Review Board of the Vaccine Research Center. Age, sex, and CMV serostatus are described in Table I. A total of four donors was studied.

Isolation of cell subsets

PBMCs were obtained by apheresis and enriched for CD8+ T cells using an Ab-depletion mixture containing reagents specific for CD4, CD14, CD19, CD36, CD66b, and CD255a (Stemcell Technologies, Vancouver, British Columbia, Canada). The CD8+ T cell-enriched sample was stained with CD36, CD66b, and CD235a (Stemcell Technologies, Vancouver, British Columbia, Canada). The CD8+ T cell-enriched sample contained from 1.95×10^6 TCR Vβ12+ cells per million (m) samples. Data were analyzed using FlowJo v9.0.1 (Tree Star). Each isotype was addressed the following hypotheses: that TCRδ clonotype frequencies will be present more frequently in the naive T cell repertoire; and that convergent recombination is a fundamental determinant of TCRβ clonotype sharing between individuals within epitope-specific memory CD8+ T cell repertoires and that they will be present more frequently in the naive T cell repertoires within and among individuals (18).

The T cell repertoire can now be probed with sufficient depth using new sequencing technology (22–24) and, when combined with high-definition polychromatic flow cytometric sorting (25), one may interrogate the clonotypic composition of stringently defined memory and naive T cell populations. In this study, we addressed the following hypotheses: that TCRδ clonotype frequency in the memory and naive CD8+ T cell pools is highly influenced by production frequency as predicted by convergent recombination; that the hierarchy of TCRβ clonotype frequencies established by V(D)J gene recombination is maintained through pairing with TCRα-chains, thymic selection, and peripheral expansion into the naive pool and subsequently into the memory T cell repertoire; and that convergent recombination is a fundamental determinant of TCRβ clonotype sharing between individuals and between the memory and naive T cell pools within individuals.

Materials and Methods

Subjects

Donors were healthy HLA-A*0201+ blood bank volunteers, who provided consent in accordance with the Institution Review Board of the Vaccine Research Center. Age, sex, and CMV serostatus are described in Table I. A total of four donors was studied.

Isolation of cell subsets

PBMCs were obtained by apheresis and enriched for CD8+ T cells using an Ab-depletion mixture containing reagents specific for CD4, CD14, CD19, CD36, CD66b, and CD255a (Stemcell Technologies, Vancouver, British Columbia, Canada). The CD8+ T cell-enriched sample was stained with CD36, CD66b, and CD235a (Stemcell Technologies, Vancouver, British Columbia, Canada). The CD8+ T cell-enriched sample contained from 1.95×10^6 TCR Vβ12+ cells per million (m) samples. Data were analyzed using FlowJo v9.0.1 (Tree Star). Each isotype was addressed the following hypotheses: that TCRδ clonotype frequency in the memory and naive CD8+ T cell pools is highly influenced by production frequency as predicted by convergent recombination; that the hierarchy of TCRβ clonotype frequencies established by V(D)J gene recombination is maintained through pairing with TCRα-chains, thymic selection, and peripheral expansion into the naive pool and subsequently into the memory T cell repertoire; and that convergent recombination is a fundamental determinant of TCRβ clonotype sharing between individuals and between the memory and naive T cell pools within individuals.

Sequence alignment and CDR3 identification

Analysis of Ag-specific CD8 T cell populations

CMV pp65 NV9-specific CD8+ T cells were stained with an allophycocyanin-labeled MHC class I tetramer and sorted by flow cytometry. Molecular analysis of all expressed TCR locus (TRB) gene products was conducted using a template-switch anchored RT-PCR, as described previously (28).

Statistical analysis

Statistical analyses were conducted using GraphPad Prism software (GraphPad) and R v2.8.1 (R Development Core 2008). Correlations were performed using the Spearman Rank test, and comparisons between two groups of data were performed using the Mann–Whitney U test.

Results

Deep sequencing enables detailed repertoire analysis of stringently defined T cell subsets

We investigated the TCRβ repertoires of precisely defined memory and naive CD8+ T cell populations from four healthy HLA-A*0201+ donors, focusing on the specific portions of the TCR repertoire defined by the single gene combination TRBV12-4/ TRBJ1-2. This approach enabled sufficient sequencing depth to facilitate a detailed comparative analysis of TCRβ repertoire composition. Importantly, potential biases arising from disproportionate usage of the many possible TRBV/ TRBJ gene combinations within the overall TCR repertoire were eliminated with this strategy. The TRBV12-4/ TRBJ1-2 gene combination was identified in previous studies as being used by the most shared human TCRB clonotype in CD8+ T cell responses to the CMV-NV9 epitope (10, 21). However, this choice of TRBV/ TRBJ1 gene combination should not influence the hypotheses being tested in this study, because the most shared TRBV12-4/TRBJ1-2 clonotype was only observed in 8 of 23 donors (21), and the TRBV12-4/TRBJ1-2 gene combination did not dominate many of the CMV-NV9–specific TCRB repertoires previously studied (10). Moreover, to determine whether CMV serostatus was an influencing factor, a CMV-seronegative donor (Donor 2) was included in this study (Table I). The memory and naive compartments of CD8+ T cell populations were isolated using polyacrylam flow cytometry.
More than 12,788 and 14,934 unique TCR sequences, we investigated the features of the unique TCR contributed to each naive repertoire (Table I). The memory T cell pool consists of clonotypes that have previously expanded in response to many different Ags encountered during an individual’s lifetime. Thus, we would expect clonotype representation in the memory pool to differ from the corresponding size distribution in the naive pool.

### Table I. Summary of TRBV12-4/TRBJ1-2 repertoire data obtained from four donors

<table>
<thead>
<tr>
<th>Donor</th>
<th>Age (y)</th>
<th>Sex</th>
<th>CMV serostatus</th>
<th>TCR sequences (n)</th>
<th>Unique a.a. clonotypes (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>Male</td>
<td>Positive</td>
<td>160,014</td>
<td>1,431</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>Male</td>
<td>Positive</td>
<td>266,362</td>
<td>34,377</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>Male</td>
<td>Negative</td>
<td>132,330</td>
<td>2,961</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>Male</td>
<td>Positive</td>
<td>277,114</td>
<td>20,422</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Donor</th>
<th>Memory</th>
<th>Naive</th>
<th>Memory</th>
<th>Naive</th>
<th>Memory</th>
<th>Naive</th>
<th>Memory</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
</tr>
<tr>
<td>2</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
</tr>
<tr>
<td>3</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
</tr>
<tr>
<td>4</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
<td>Memory</td>
<td>Naive</td>
</tr>
</tbody>
</table>

The memory T cell pool consists of clonotypes that have previously expanded in response to many different Ags encountered during an individual’s lifetime. Thus, we would expect clonotype representation in the memory pool to differ from the corresponding size distribution in the naive pool. Indeed, the distributions of amino acid clonotype sizes differed substantially between the memory and naive TCRβ repertoires in all four donors, with greater skewing in the memory pool (Fig. 2A, Supplemental Fig. 1A–F). For example, the 10% of memory TCRβ amino acid clonotypes with the largest sizes in Donor 1 contributed 96.5% of the memory repertoire; in contrast, the 10% of naive TCRβ amino acid clonotypes with the largest sizes contributed only 46.7% of the naive repertoire. Thus, the unevenness of memory TCRβ amino acid clonotype size distributions was largely due to several highly dominant clonotypes (Fig. 2B).

### FIGURE 1. Isolation of memory and naive CD8+ T cells by high-definition flow cytometric sorting.

Sorting scheme for the isolation of memory and naive CD8+ T cells. These cells were subsequently analyzed on the basis of multiple cell surface marker expression (middle and bottom rows). Naive cells were defined as CD27-CD45RO-CD57-CD127CCR7+, and memory cells were defined as CD27+CD45RO+, CD27+CD45RO*, and CD27+CD45RO-CD57+. Both subsets expressed TCR Vβ12.

### FIGURE 2. Clonotype size distributions differ between the memory and naive TCRβ repertoires.

The memory T cell pool consists of clonotypes that have previously expanded in response to many different Ags encountered during an individual’s lifetime. Thus, we would expect clonotype representation in the memory pool to differ from the corresponding size distribution in the naive pool. Indeed, the distributions of amino acid clonotype sizes differed substantially between the memory and naive TCRβ repertoires in all four donors, with greater skewing in the memory pool (Fig. 2A, Supplemental Fig. 1A–F). For example, the 10% of memory TCRβ amino acid clonotypes with the largest sizes in Donor 1 contributed 96.5% of the memory repertoire; in contrast, the 10% of naive TCRβ amino acid clonotypes with the largest sizes contributed only 46.7% of the naive repertoire. Thus, the unevenness of memory TCRβ amino acid clonotype size distributions was largely due to several highly dominant clonotypes (Fig. 2B). Individuation differences in TCRβ amino acid clonotype size distributions were also observed, albeit on a smaller scale compared with the differences between the memory and naive pools. In particular, the sizes of naive TCRβ amino acid clonotypes were less evenly distributed in Donor 4, and the sizes of memory TCRβ amino acid clonotypes were more evenly distributed in Donor 2, compared with the other three donors (Supplemental Fig. 1A, 1B).

The most dominant memory TCRβ clonotypes made up 61, 35, 20, and 24% of the memory TCRβ repertoires of Donors 1, 2, 3, and 4, respectively. We investigated whether persistent CMV infection could potentially play a role in the surprising dominance of the memory TCRβ clonotypes in the three CMV-seropositive donors. A search of the memory TCRβ repertoires of Donors 1, 3, and 4 for TCRβ clonotypes specific for the immunodominant CMV-NV9 epitope from previous studies (10, 21, 30–34) revealed only three previously identified CMV-NV9–specific TCRβ clonotypes. All three of these previously identified CMV-NV9–specific TCRβ clonotypes were found in Donor 1 and included the second most dominant memory TCRβ clonotype, CASSLVG-GRYGTF. We determined that the most dominant memory TCRβ clonotype in Donor 1, CASAYGAYGNYTF, was also specific for the CMV-NV9 epitope by sequencing the CMV-NV9–specific TCRβ repertoire of Donor 1 using conventional Sanger sequencing. Although CMV-specific T cell responses seem to play a role in the dominance of memory TRBV12-4/TRBJ1-2 clonotypes in Donor 1, the prevalence of the most dominant memory TCRβ clonotype in CMV-seronegative Donor 2, which was comparable to that of the most dominant memory TCRβ clonotypes in CMV-seropositive Donors 3 and 4, suggests that persistent CMV infection is not the
FIGURE 2. Composition of the memory and naive TCRβ amino acid repertoires. A. Distributions of clonotype sizes in the memory and naive repertoires of Donor 1 represented as the cumulative percentage of the total TCRβ amino acid (a.a.) repertoire versus the cumulative percent of unique TCRβ amino acid clonotypes ordered by increasing clonotype size. In this representation, an even clonotype size distribution (i.e., all clonotypes being the same size) would be indicated by the gray dashed line, and the extent of deviation from this line indicates the unevenness of the distribution. This plot shows, for example, that the 10% of naive TCRβ amino acid clonotypes with the largest sizes contribute 46.7% of the total naive TCRβ repertoire; in contrast, the 10% of memory TCRβ clonotypes with the largest sizes contribute 96.5% of the total memory TCRβ repertoire. B. Individual dominant TCRβ amino acid clonotypes in the memory (left panel) and naive (right panel) repertoires of Donor 1. Clonotypes that make up ≥0.5% of the total TCRβ amino acid repertoire are shown as color-shaded segments. The TCRβ amino acid clonotypes are represented by the CDR3 amino acid sequence between the conserved cysteine and phenylalanine. The light and dark gray segments are composed of a number of TCRβ amino acid clonotypes, as indicated on the plots. Segments shaded in dark gray correspond to TCRβ clonotypes making up <0.5% and ≥0.1% of the total amino acid repertoire. Segments shaded in light gray correspond to TCRβ sequences making up <0.1% of the total amino acid repertoire. C. Distribution of CDR3 lengths among the unique TCRβ amino acid clonotypes in the memory and naive repertoires of Donor 1. D. Distribution of CDR3 lengths for the total TCRβ amino acid memory and naive repertoires (i.e., accounting for clonotype size) of Donor 1. CDR3 length is the number of residues between the conserved cysteine and phenylalanine residues, as defined by IMGT (26).

FIGURE 3. TCRβ amino acid clonotypes common to the memory and naive repertoires. A. Left panel. For each of the four donors, the proportion of unique memory and naive TCRβ amino acid (a.a.) clonotypes that are common to the memory and naive pools. Right panel. For each of the four donors, the proportion of the total memory and naive TCRβ repertoires attributable to TCRβ amino acid clonotypes that are common to the memory and naive pools. B. Left panel. Comparison of TCRβ amino acid clonotype sizes in the memory repertoire of Donor 1 between clonotypes unique to the memory pool and those also present in the naive pool. Right panel. Comparison of the TCRβ amino acid clonotype sizes in the naive repertoire of Donor 1 between clonotypes unique to the naive pool and those also present in the memory pool. Statistics are based on the Mann–Whitney U test. C. Relationship between TCRβ amino acid clonotype sizes in the naive and memory repertoires for clonotypes common to both pools in Donor 1. The opacity of the circles indicates the density of points plotted. Correlations are based on the Spearman rank test.
sole determinant of the surprising dominance of these memory clonotypes.

**CDR3 length distributions differ between the memory and naive TCRβ repertoires**

The distribution of CDR3 sequence lengths is another feature that provides an overall view of repertoire composition. Biases in CDR3 length are often observed in epitope-specific T cell repertoires, suggesting that CDR3 length distributions might differ substantially between the memory and naive T cell pools. Surprisingly, we found that the distributions of CDR3 lengths among unique TCRβ amino acid clonotypes were similar in the memory and naive pools within individuals and between individuals. The median CDR3 length was 12 aa in the memory and naive pools of all donors (Fig. 2C, Supplemental Fig. 1G–J). However, this parameter provides little indication of the CDR3 lengths that predominate within the total repertoires. Therefore, we also assessed the distributions of CDR3 lengths across the total memory and naive TCRβ amino acid repertoires (i.e., including the size of each clonotype). We found substantial CDR3 length distribution differences between the memory and naive pools, largely due to the prevalence of one or two CDR3 lengths in the memory TCRβ repertoires of each donor (Fig. 2D, Supplemental Fig. 1K–N). The preponderance of these few CDR3 lengths in the memory repertoires was largely associated with the most dominant memory TCRβ clonotypes. For example, the dominance of the CASALGNYGYTF clonotype in the memory repertoire of Donor 1 (Fig. 2B) was largely responsible for the peak at 11 aa (Fig. 2D). There were also notable interindividual differences in CDR3 length distributions among all memory TCRβ amino acid sequences between the four donors (Supplemental Fig. 1K–N). Thus, although the memory and naive repertoires consisted of TCRβ amino acid clonotypes with similar CDR3 length features, most likely owing to the large variety of Ag specificities of T cells in the memory pool, differences in the most prevalent CDR3 lengths were observed between the memory and naive pools and between individuals.

**Dominant memory TCRβ clonotypes are highly represented in the naive pool**

Next, we examined the extent of overlap of TCRβ amino acid clonotypes between the memory and naive CD8+ T cell compartments. In each donor, a subset of TCRβ amino acid clonotypes was common to the memory and naive pools. Of the unique TCRβ amino acid clonotypes in the memory pool of a donor, a mean of 10.2% (range: 3.7–15.3%) was also present in the naive pool (Fig. 3A). A much smaller percentage of the TCRβ amino acid clonotypes in the naive repertoires was observed in the memory pools (mean, 1.1%; range: 0.6–1.4%). We then determined the contribution that these clonotypes present in memory and naive pools made to the total memory and total naive TCRβ repertoires (i.e., the extent of overlap of the total memory and total naive repertoires). Despite large interindividual variations, a substantial proportion of each donor’s total memory TCRβ repertoire consisted of amino acid clonotypes that were also present in the naive pool (mean, 69.1%; range: 48.8–92.5%; Fig. 3A). In contrast, a much smaller proportion of an individual’s total naive TCRβ repertoire overlapped with the memory pool (mean, 3.5%; range: 2.6–4.2%).

To gain a better understanding of the overlap between the memory and naive repertoires, we also investigated the commonality of TCRβ clonotypes at the nucleotide level between the

---

**FIGURE 4.** Convergent recombination. Using examples of TCRβ clonotypes prevalent in the TRBV12-4/TRBJ1-2 repertoires, the various levels of convergent recombination are illustrated. **A.** At the level of the amino acid sequence, multiple nucleotide sequences converge to code for the same TCRβ amino acid clonotype. The TCRβ clonotype, CASSALGNYGYTF, was observed in all four donors, and it was common between the naive and memory compartments in two of these donors. This clonotype was found to be encoded by a variety of nucleotide sequences across all donors’ repertoires. One possible recombination mechanism is shown for each nucleotide sequence involving the TRBV12-4 (blue), TRBD1 (pink), and TRBJ1-2 (green) genes and a minimal number of nucleotide additions (black). **B.** At the level of the nucleotide sequence, multiple recombination mechanisms, involving different gene contributions and nucleotide additions, converge to produce the same nucleotide sequence. This is shown for one of the nucleotide sequences (red box) encoding CASSALGNYGYTF using a variety of recombination mechanisms requiring similarly low numbers of nucleotide additions. This particular nucleotide sequence was shared among three of the four donors. **C.** The relevant human TRB gene sequences.
memory and naive compartments of a donor’s repertoire. A large proportion of the total memory repertoire (mean, 62.9%; range: 33.8–90.4%) consisted of TRβ nucleotide clonotypes that were also present in the naive pool. One explanation of this overlap is simply contamination of naive cells in the memory repertoire or vice versa. However, more than half (mean, 56.8%) of the TRβ amino acid clonotypes common to memory and naive pools were encoded by at least one nucleotide sequence that was present in only one of the compartments. This suggests that, although there is substantial overlap between the memory and naive pools at the nucleotide sequence level, this only partially contributes to the substantial overlap observed at the TRβ amino acid sequence level.

To assess quantitatively the dominance of the TRβ amino acid clonotypes common to the memory and naive pools within an individual, we compared the size of these clonotypes with those that were unique to one of the compartments. In the memory repertoires, TRβ amino acid clonotypes unique to the memory pool had a significantly smaller size compared with those that were also found in the naive pool (Fig. 3B, Supplemental Fig. 2A–D). Similarly, in the naive repertoires, the size of TRβ amino acid clonotypes unique to the naive pool was significantly smaller compared with those also present in the memory pool (Fig. 3B, Supplemental Fig. 2E–H). Furthermore, among the TRβ amino acid clonotypes common to the memory and naive pools within a donor, we observed a positive correlation between clonotype size in the naive pool and clonotype size in the memory pool (Fig. 3C, Supplemental Fig. 2I–L). Thus, a substantial portion of the memory TRβ repertoire at the amino acid level overlapped with the naive TRβ repertoire within each donor. The extent of the overlap was largely due to the dominance in the memory pool of clonotypes also present in the naive pool.

Convergent recombination is a determinant of relative TRβ clonotype size

To establish a mechanistic basis for the observed TRβ clonotype dominance hierarchies, we investigated whether the process of convergent recombination could predict relative clonotype sizes in the memory and naive CD8+ T cell pools. Our previous studies of epitope-specific memory repertoires showed that convergent recombination enables particular TRβ clonotypes to be produced more easily than others during gene rearrangement (18–21). This process operates at two levels: particular TCR nucleotide sequences can be produced more efficiently by one or more frequently occurring V(D)J recombination events (35) and/or by many closely related recombination events (i.e., different contributions of the gene segments and nucleotide additions) (18–21) and particular TCR amino acid sequences can be easily made if they can be encoded by nucleotide sequences that are efficiently produced and/or encoded by many different nucleotide sequences. The latter is largely determined by codon degeneracy of specific amino acids in the CDR3 sequence. The convergent recombination process is demonstrated in Fig. 4 using, as an example, a TRBV12-4/TRBJ1-2 amino acid clonotype that was observed in all four individuals. A resulting prediction is that the frequencies at which TCR sequences are produced by V(D)J recombination should shape clonotype dominance hierarchies in the naive T cell repertoire.

A key indicator that a TCR nucleotide sequence has the potential to be produced efficiently by V(D)J recombination is that it requires fewer nucleotide additions and, thus, comprises less of a random element. To determine whether the relative TRβ clonotype sizes in the peripheral repertoires were associated with the number of nucleotide additions, we calculated the minimal number of nucleotide additions required to produce each of the TRβ sequences by sequentially aligning the TRBV12-4, TRBJ1-2, and TRBD genes (Fig. 5A, Supplemental Fig. 3A–D). For the memory and naive CD8+ T cell populations in each donor, the estimated minimal number of nucleotide additions required to produce a TRβ nucleotide clonotype was negatively correlated with the size of the TRβ nucleotide clonotype in the repertoire (Fig. 5C, Supplemental Fig. 3E–L). Comparisons between the TRβ amino acid clonotypes unique to memory pool and those also present in the naive pool and between the TRβ amino acid clonotypes unique to naive pool and those also present in the memory pool revealed that nucleotide sequences coding for TRβ amino acid clonotypes are encoded by at least one nucleotide sequence that was present in only one of the compartments. This suggests that, although there is substantial overlap between the memory and naive pools at the nucleotide sequence level, this only partially contributes to the substantial overlap observed at the TRβ amino acid sequence level.

![Figure 5](http://www.jimmunol.org/) Relationship between convergent recombination and TRβ clonotype size in the memory and naive repertoires. A. Distribution of the minimal number of nucleotide (n.t.) additions required for the production of memory and naive TRβ clonotypes from Donor 1. B. Distribution of the number of different nucleotide sequences encoding memory and naive TRβ amino acid (a.a.) clonotypes from Donor 1. The percentages of unique TRβ amino acid clonotypes at which the distributions plateau at higher numbers of encoding nucleotide sequences correspond to one TRβ amino acid clonotype. C. Relationship between the minimal number of nucleotide additions and TRβ nucleotide clonotype size in the memory (left) and naive (right) repertoires of Donor 1. D. Relationship between the number of different nucleotide sequences observed to encode a TRβ amino acid clonotype and its size in the memory (left) and naive (right) repertoires of Donor 1. TRβ clonotype sizes are grouped according to the nonoverlapping categories on the horizontal axis. For example, ≤10^2 refers to a number of copies >10 and ≤10^2. Correlations are based on the Spearman rank test applied to the noncategorized data. In C and D, the total numbers of unique TRβ clonotypes are indicated for each group of data above the box representing the data group.
clonotypes common to the memory and naive pools required significantly fewer nucleotide additions (p < 0.01 for each comparison for Donors 1–4; Mann–Whitney U test).

We also investigated whether the relative sizes of memory and naive TCRβ amino acid clonotypes were associated with the variety of encoding nucleotide sequences, which is another key indicator of TCRβ amino acid sequence production efficiency. The distributions of the number of nucleotide sequences observed to encode each memory and naive TCRβ amino acid clonotype are shown in Fig. 5B and Supplemental Fig. 3M–P. In all four donors, the number of different encoding nucleotide sequences was positively correlated with TCRβ amino acid clonotype size in the memory and naive repertoires (Fig. 5D, Supplemental Fig. 3Q–X). Comparisons between the TCRβ amino acid clonotypes unique to memory pool and those also present in the naive pool and between the TCRβ amino acid clonotypes unique to naive pool and those also present in the memory pool revealed that TCRβ amino acid clonotypes common to the memory and naive pools were encoded by a greater variety of nucleotide sequences (p < 0.0001 for each comparison for Donors 1–4; Mann–Whitney U test).

Thus, TCRβ clonotypes with the potential to be produced more efficiently during gene recombination are more likely to be present at a higher frequency in the naive repertoire. Remarkably, this also holds for the memory repertoire, suggesting that the dominance hierarchy established by convergent recombination during TCR production is, to some extent, preserved through the generation of memory populations.

**Clonotype size is a determinant of interindividual TCRβ clonotype sharing**

The many previous descriptions of interindividual sharing of identical epitope-specific clonotypes (17) imply that there must be a degree of TCR repertoire overlap between different individuals. However, the extent of overlap between the total peripheral repertoires is not known. Therefore, we initially determined the TCRβ amino acid clonotypes that were shared between each of the six possible pairings among the four donors (i.e., Donors 1 and 2, Donors 1 and 3, Donors 1 and 4, Donors 2 and 3, Donors 2 and 4, and Donors 3 and 4). We considered two measures of the extent of TCRβ sharing between pairs of donor. First, we evaluated the proportion of unique TCRβ clonotypes across the memory and naive pools of a donor that were shared with each of the other three donors. Between 3.8 and 9.8% of the unique TCRβ amino acid clonotypes across the memory or naive CD8+ T cell populations of a donor were shared with another donor (Fig. 6A). Second, as a measure of the interindividual overlap between total repertoires (i.e., allowing for clonotype sizes), we considered the proportions of the total memory and total naive repertoires in a donor that were attributable to TCRβ amino acid clonotypes shared with each of the other three donors (Fig. 6B, 6C). The proportion of a donor’s total naive repertoire that was contributed by TCRβ amino acid clonotypes present in each of the other three donors varied between 2.1 and 83.9%, with a mean across all six pairwise comparisons between donors of 37.5% (Fig. 6B).

To examine the extent of sharing of TCRβ amino acid clonotypes across all four donors, we determined the number of individuals in which each TCRβ amino acid clonotype was present. Identical TCRβ amino acid clonotypes could be found in two, three, or all four individuals (Fig. 7). As with the comparisons between pairs of donors above, we considered two measures of the extent of TCRβ sharing across the four donors: the proportion of unique TCRβ clonotypes in the memory and naive pools that were shared with other individuals (Fig. 7A) and the proportions of the total memory and total naive repertoires that were attributable to shared TCRβ amino acid clonotypes (Fig. 7B). Notably, TCRβ amino acid clonotype size, in the memory and naive pools of an individual, was positively correlated with the number of individuals in which that TCRβ clonotype was observed (Fig. 7C, Supplemental Fig. 4). Thus, TCRβ amino acid clonotypes that are prevalent in individual memory and naive repertoires are more likely to be observed in many individuals.

**Many TCRβ clonotypes common to the memory and naive pools within individuals are also shared between individuals**

Next, we investigated the extent of interindividual sharing of TCRβ clonotypes common to the memory and naive repertoires within individuals. The overlapping portions of a donor’s memory

**FIGURE 6.** Sharing between pairs of donors of TCRβ amino acid clonotypes in the memory and naive repertoires. The TCRβ amino acid (a.a.) clonotypes shared between each pair of donors was evaluated, regardless of whether they were found in the memory and/or naive pool. A. The proportion of unique TCRβ amino acid clonotypes in Donors 1–4 (i.e., pooled across the memory and naive repertoires) that were observed in each of the other three donors. The proportions of the total memory (B) and naive (C) TCRβ repertoires of Donors 1–4 that were attributable to TCRβ amino acid clonotypes observed in each of the other three donors. For example, B shows that 83.9% of the memory TCRβ repertoire in Donor 4 was made up of TCRβ amino acid clonotypes that were also observed in Donor 1.
and naive pools consisted of a mean of 56.7% (range: 51.1–66.3%) unique TCRβ amino acid clonotypes that were also shared between individuals (Fig. 7D). Indeed, a substantial proportion of TCRβ amino acid clonotypes present in the memory and naive pools within a donor were also found in more than two individuals. Thus, a higher degree of TCRβ amino acid clonotype sharing was observed among clonotypes common to the memory and naive pools of an individual than among those unique to either pool.

Discussion

Although recent studies highlighted the importance of T cell repertoire composition for immune recognition of specific pathogens (3, 36–38) and the maintenance of immune efficacy with age (39), our current understanding of these processes is derived predominantly from analyses of small samples of epitope-specific repertoires. Next-generation sequencing technologies provide an opportunity to study complete memory and naive T cell repertoires in depth. In this study, we rigorously sorted the memory and naive CD8+ T cell populations from four donors and pyrosequenced the portions of the corresponding TCRβ repertoires with TRBV12-4/TRBJ1-2 gene rearrangements. We found that there is a high degree of overlap between the memory and naive repertoires within individuals; TCRβ clonotypes common to the memory and naive pools tend to be more frequent in both pools; a substantial proportion of the memory and naive clonotypes are indicated for each group of data above the bar or box representing the data group.

The presence of particular TCRβ clonotypes in the memory and naive pools evokes several interesting interpretations. One scenario is that naive T cells bearing the same TCR are entirely recruited into the memory pool during an Ag-specific response. The presence of those same TCRβ clonotypes in the naive pool would then derive from thymic replenishment of the naive repertoire with identical TCR sequences. Indeed, recent studies in mice showed that, although governed by affinity for peptide–MHC (11, 41), recruitment of Ag-specific CD8+ T cells from the naive pool during Ag encounter (7, 8, 40). Although there were substantial differences in clonotype size and CDR3 length distribution between the memory and naive pools, a subset of TCRβ clonotypes was common to both pools. Many of these TCRβ clonotypes were relatively dominant in the memory and naive pools, with clonotype size in the memory pool being significantly associated with clonotype size in the naive pool. However, the more pronounced dominance hierarchy among memory TCRβ clonotypes compared with naive TCRβ clonotypes resulted in a surprisingly high degree of overlap within the total memory repertoire with the total naive repertoire among individuals.

The high degree of interindividual TCRβ clonotype sharing between the memory and naive clonotypes strongly suggests that these TCRβ clonotypes were initially present at a relatively high frequency in the naive TCR repertoires of each of the donors. Alternatively, it might be that not all cells with a particular TCRβ clonotype are recruited from the naive pool by a particular Ag. This could occur as a result of suboptimal clonotypic TCR affinities for peptide–MHC, differential TCRα/chain pairing with altered fine specificity, or simply a lack of Ag encounter. Another interpretation is that asymmetric division and differentiation after Ag encounter result in T cells bearing the same TCRβ clonotypes having memory and naive phenotypes. Such a model has been proposed in mice (42) and is supported by the overlap between memory and naive...
repertoires at the level of nucleotide sequence. However, the observation that different nucleotide sequences in the memory and naive repertoires also encoded many common TCRβ amino acid clonotypes makes it quite clear that these common TCRβ clonotypes originated from different cells, which runs somewhat contrary to this explanation. Importantly, these various interpretations are not mutually exclusive and may all contribute to TCRβ clonotype sharing between the memory and naive pools.

TCRβ clonotypes that are shared between individuals are thought to play an important role in the efficacy of pathogen-specific responses and the control of infection (3, 36). Thus, an understanding of the mechanisms that determine interindividual TCR sharing in humans is likely salient to vaccine development. We found that a substantial proportion (>24.5%) of the naive TCRβ amino acid repertoire within an individual was shared with at least one of the other donors in this study. Thus, the extent of TCRβ sharing between larger groups of individuals should be potentially much greater. Indeed, studies in syngeneic mice showed that up to 27% of the peripheral repertoire of one naive mouse overlap with that of another (16). Our previous studies of epitope-specific TCRβ repertoires suggested that a process of convergent recombination enables some TCRβ clonotypes to be produced by V(D)J recombination in the thymus more frequently than others. Such TCRβ clonotypes are predicted to be present at a greater frequency in the naive repertoire and have a greater likelihood of being shared between individuals (18–21). Our investigations of the relationship between the sharing of TCRβ clonotypes between the donors in this study and clonotype sizes in the memory and naive pools provide strong evidence for convergent recombination as a molecular basis for interindividual TCR sharing.

We also determined to what extent the hierarchy of TCRβ production frequency predicted by convergent recombination is modified by other processes, such as TCRβ production frequency predicted by convergent recombination is a fundamental determinant of clonotype size in the memory and naive T cell pools. Furthermore, TCRβ clonotype size influences the overlap between the memory and naive TCRβ repertoires, as well as interindividual clonotype sharing. Interestingly, outliers in the data suggest the involvement of other factors in shaping the repertoire. That is, not all TCRβ clonotypes predicted to be easily generated by convergent recombination were present in large numbers in the memory and/or naive repertoires of multiple individuals. The variety of TCRα-chains with which the TCRβ-chains pair, the efficiency with which these TCRα-chains are made, the proportion of these TCRβ heterodimers that survive thymic selection, and the efficiency of their recruitment into the memory pool are all important factors that determine the presence and size of TCRβ clonotypes in the memory and naive pools. Nonetheless, the associations identified in this study provide new insights into the mechanisms that shape the peripheral T cell repertoires and reveal the profound influence of differential clonotype production frequencies in the thymus as a consequence of the process of convergent recombination. This mechanistic understanding may prove invaluable for the design of effective T cell-based vaccines that aim to exploit established correlates of immune control (3, 43).

Acknowledgments
We thank D. Ambrozak, R. Nguyen, and S. Perfetto for expert assistance with flow cytometry, and M. Roederer and J. Yu of the Vaccine Research Center Flow Cytometry Core for production and conjugation of Abs.

Disclosures
The authors have no financial conflicts of interest.

References
15. Baron, V., C. Bouneaud, A. Cumano, A. Lim, T. P. Arstila, P. Kourilsky, L. Ferradini, and C. Pannetier. 2003. The repertoires of circulating human CD8...


