

Very Diverse CD8 T Cell Clonotypic Responses after Virus Infections¹

Lecia L. Pewe,* Jason M. Netland,* Stephen B. Heard,[‡] and Stanley Perlman^{2*†}

We measured CD8 T cell clonotypic diversity to three epitopes recognized in C57BL/6 mice infected with mouse hepatitis virus, strain JHM, or lymphocytic choriomeningitis virus. We isolated epitope-specific T cells with an IFN- γ capture assay or MHC class I/peptide tetramers and identified different clonotypes by V β chain sequence analysis. In agreement with our previous results, the number of different clonotypes responding to all three epitopes fit a log-series distribution. From these distributions, we estimated that >1000 different clonotypes responded to each immunodominant CD8 T cell epitope; the response to a subdominant CD8 T cell epitope was modestly less diverse. These results suggest that T cell response diversity is greater by 1–2 orders of magnitude than predicted previously. *The Journal of Immunology*, 2004, 172: 3151–3156.

The adaptive immune system must be able to recognize a large repertoire of foreign Ags, while including enough naive precursor cells to any Ag to generate an infection-controlling response. The diversity of the $\alpha\beta$ TCR response results from imprecise recombination of the V, D, and J regions of α - and β -chains coupled with pairing of one β -chain and one to two α -chains per T cell (1). The theoretical diversity of the T cell repertoire is $\sim 10^{15}$ (2), but the total number of CD8 T cells in a single mouse is estimated to be 2×10^7 , with a diversity of 2–5 $\times 10^5$ different clonotypes per mouse spleen (3, 4).

The number of precursor cells specific to a single epitope appears to be too low to estimate via direct isolation with current techniques. Several alternative approaches to this problem have yielded very different values for T cell clonotype diversity. For instance, Maryanski et al. and Bousso et al. (5, 6) exploited the limited CD8 T cell response to HLA-A2 and HLA-CW3 in DBA/2 mice (H-2K^d restricted). CD8 T cells responding to single peptides within these Ags primarily use the V β 10 element with a complementarity-determining region 3 (CDR3)³ length of 6 or 10 aa. By isolating and sequencing TCR cDNA clones corresponding to this subgroup, the authors estimated that ~ 200 cells or 15–20 different clonotypes per spleen recognized this epitope. In another approach, the precursor frequency to the H-2D^b-restricted gp33–41 epitope of lymphocytic choriomeningitis virus (LCMV, epitope gp33) was estimated from the number of transferred LCMV-specific TCR transgenic CD8 T cells required to give a response equivalent to the endogenous response (4). The authors calculated

that there were ~ 100 –200 CD8 T cells or 20 different clonotypes specific for epitope gp33 in the naive adult C57BL/6 (B6) mouse.

Recently, we used an alternative approach to determine the precursor frequency to a well-defined, immunodominant CD8 T cell epitope recognized in mice infected with mouse hepatitis virus (strain JHM) (7). Intranasal infection with this virus causes acute encephalitis with a robust T cell response (8). The CD8 T cell response in B6 mice is directed against an immunodominant and a subdominant epitope, encompassing residues 510–518 (epitope S510, H-2D^b restricted) and residues 598–605 (epitope S598, H-2K^b restricted) of the spike (S) glycoprotein (9, 10). To determine the number of epitope S510-specific CD8 T cells in the infected CNS, we isolated epitope-specific T cells with MHC class I/peptide tetramers and sequenced CDR3 regions from individual cDNA clones. The frequency of epitope S510-positive V β 13-expressing clones fits a log-series distribution. We estimated that there were 300–500 epitope S510-specific CD8 T cell clonotypes present in the acutely infected CNS. This number is higher than estimated previously (4, 5), but is still probably an underestimate, because it does not consider redundancy in the T cell population or the contribution of the α -chain to TCR diversity.

To address the generality of our T cell diversity estimates and to examine the response in a secondary lymphoid tissue, we determined the number of clonotypes responding to epitopes S510 and S598 in JHM-infected mice and to epitope gp33 in LCMV-infected mice, after i.p. inoculation.

Materials and Methods

Mice

Specific pathogen-free B6 mice were purchased from the National Cancer Institute (Bethesda, MD). Animal studies were approved by the University of Iowa Animal Care and Use Committee.

Virus

JHM was grown and titered, as described (11). Mice were infected i.p. with 1.5×10^5 PFU JHM or 5×10^5 PFU Armstrong strain of LCMV in PBS.

IFN- γ capture assay

Cells were harvested from spleens of JHM-infected mice 7 days after infection. To isolate epitope S510- or S598-specific CD8 T cells, 1×10^7 unfractionated splenocytes/ml were stimulated 1:50 with EL-4 cells for 4 h with 1 μ M of S510 peptide (CSLWNGPHL) or S598 peptide (RCQIFANI) (12). After stimulation, epitope-specific CD8 T cells were captured using the MACS mouse IFN- γ secretion assay (Miltenyi Biotec, Auburn, CA) following the manufacturer's protocol for frequencies of IFN- γ -secreting T

Departments of *Pediatrics and [†]Microbiology, University of Iowa, Iowa City, IA 52242; and [‡]Department of Biology, University of New Brunswick, Fredericton, New Brunswick, Canada

Received for publication September 4, 2003. Accepted for publication December 30, 2003.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ This research was supported by grants from the National Institutes of Health (NS40438) and National Multiple Sclerosis Society (RG 2864-B-3).

² Address correspondence and reprint requests to Dr. Stanley Perlman, Department of Pediatrics, University of Iowa, Medical Laboratories 2042, Iowa City, IA 52242. E-mail address: stanley-perlman@uiowa.edu

³ Abbreviations used in this paper: CDR3, complementarity-determining region 3; JHM, mouse hepatitis virus; LCMV, lymphocytic choriomeningitis virus; S, surface glycoprotein.

cells <2%. Cells were additionally stained with FITC-conjugated anti-CD8 Abs. IFN- γ -expressing CD8 T cells were sorted with a FACS DiVa (BD Biosciences, San Jose, CA). A total of 50,000–220,000 epitope S510- or S598-specific cells was obtained from each sort. Under our sorting conditions, virtually no IFN- γ -expressing CD8 T cells were detected when cells were stimulated with no added peptide or when cells from uninfected mice were stimulated with JHM-specific peptide (Fig. 1).

Staining with tetramers

Splenocytes harvested from LCMV-infected mice were stained with MHC class I/gp33 peptide (KAVYNFATM) tetramer complexes (kindly provided by K. Messingham and J. Harty, University of Iowa), as described (7). A total of 100,000 tetramer gp33-positive CD8 cells was sorted with a FACS DiVa. As a negative control, cells were also exposed to irrelevant tetramer (specific for JHM epitope S510) (Fig. 1).

Isolation of RNA from lymphocytes

RNA was isolated from lymphocytes using an RNeasy Mini Kit (Qiagen, Valencia, CA).

Sequence analysis

cDNA was synthesized from each entire sorted lymphocyte RNA preparation in a 60 μ l reaction, as described (13). A total of 5 μ l of cDNA was used for PCR. For synthesis of V β 5-, V β 8-, and V β 13-specific PCR products, V β 5-specific (CCAGCAGATTCTCAGTCCAAC), V β 8-specific (CATGGGCTGAGGCTGATCCATT), and V β 13-specific (CCTAAAG GAATAACTCCACTCT) primers were used with a common C β reverse primer (GCAATCTCTGCTTTTGATGGCTC) and *Taq* polymerase (AmpliTaq Gold DNA Polymerase; Applied Biosystems, Foster City, CA). Amplification was conducted for 25 cycles, as described previously (13). The PCR error rate was 0.12% (2/1680, based on sequence analysis of conserved residues of the V β and J β chains). PCR products were cloned into plasmid vector pCR2.1-TOPO, as described by the manufacturer (Invitrogen, Carlsbad, CA). Colonies containing inserts were identified and sequenced with the V β primers above and an automated ABI Prism 3700 Sequencer (Applied Biosystems).

TCR repertoire analysis

We fit log-series distributions to our clonotype diversity data. This distribution is often used to describe data with a few abundant and many rare species (14, 15). A log-series distribution has the form $s_i = \alpha x^i / i$, where s_i is the number of CDR3 expected to be represented i times, α is an index of clonal diversity, and x is a parameter related to diversity and sample size (number of cDNA clones sequenced). The diversity index α depends on S , the number of different (at the nucleotide level) CDR3 species present in the T cell population, and N , the total number of T cells in the population: $S = \alpha \ln(1 + N/\alpha)$. α is large when the number of different CDR3 is high relative to the number of T cells. Our calculation of α followed *Krebs* (16) and was implemented in a computer program written in QuickBASIC 4.5 (Microsoft, Redmond, WA). Briefly, we estimated the parameter x using an iterative solution to the equation $S/N = [(1 - x)/x][-\ln(1 - x)]$, and then calculated our estimate of α as $\alpha = N(1 - x)/x$ (see Ref. 16, equations 12.12 and 12.13). The expected number of clonotypes found in a larger sample of cells (say, an entire spleen) is easily calculated: $S' = \alpha \ln(1 + N'/\alpha)$. Sample calculations are included as Appendix. SEs for α were obtained as described (17).

In a log-series distribution, the modal frequency is always 1 (that is, there are more clonotypes represented once than with any other integer frequency). However, this does not necessarily mean that most clonotypes are singletons, or even that most clonotypes are rare, because the total number of common clonotypes (summing, say, the frequencies of clonotypes represented 10 times, 11 times, 12 times. . . .) can be substantial. For instance, if $S = 10$ and $N = 100$ (10 distinct clonotypes in 100 sequenced cells), then ~ 7 of the 10 clonotypes are expected to be present at least twice (Appendix). For each mouse, goodness-of-fit of the data to a log series was assessed using a G test with Williams' correction (18), using observed and expected counts for five frequency classes (sequences represented once, 2–3 times, 4–7 times, 8–15 times, and 16 or more times; but results were robust to different lumpings). Calculation of expected frequencies is illustrated in Appendix. We similarly tested goodness-of-fit to uniform frequency distributions. To increase statistical power, we combined the test results across all data sets using Fisher's method (17) to yield a single, powerful test of each distributional hypothesis.

Results

Detection of epitope S510-specific T cells

To extend our studies to a second CD8 T cell epitope, we used an IFN- γ capture assay to isolate epitope S598-specific T cells from the spleens of mice after inoculation with JHM. Previously, we isolated epitope S510-specific CD8 T cells with MHC class I/peptide tetramers (7). However, epitope S598 had low functional avidity (amount of peptide required to sensitize target cells for lysis) (10) and consistent with this, epitope S598-specific tetramers rapidly dissociated from T cells, resulting in poor cell recovery after cell sorting. In preliminary experiments, intact RNA suitable for RT-PCR was not reproducibly isolated after detection of epitope-specific cells by conventional intracellular IFN- γ assay. The IFN- γ capture assay does not require fixation of cells, and so allows isolation of high quality RNA. We measured precursor frequencies in the spleen after i.p. infection, in part because cell recovery from the CNS was low after IFN- γ capture, and in part to document the repertoire in a secondary lymphoid organ. In initial experiments, we detected the same percentage of Ag-specific T cells with the IFN- γ capture as with intracellular IFN- γ -staining assays. We also infected mice with 5×10^4 , 1.5×10^5 , or 5×10^5 PFU of JHM and found that 1.5×10^5 PFU gave a maximal T cell response.

Initially, we measured the number of CD8 T cell clonotypes responding to epitope S510 and compared it with our previous results for cells isolated from the infected CNS. Epitope S510-specific T cells were sorted by flow cytometry and V β 13-expressing cells were analyzed further because these cells were most abundant in the epitope S510 population (7). Under our conditions of cell sorting, virtually no cells were IFN- γ^+ in the absence of peptide S510 stimulation (Fig. 1). We analyzed 52 and 58 cDNA clones from each of two mice, with 29 and 31 different CDR3 detected; 17 and 22 of these were identified in single clones (Table I). As in the CNS, the response was polyclonal with diverse J β usage, but preferential usage of J β 2.1 and J β 2.6 (Supplementary Table I)⁴ (7). Only one CDR3 was common to both mice, but 4 of 60 of these spleen-derived CDR3 were previously detected in the CNS (7) (Supplementary Table I).

As for CNS-derived cells (7, 19), frequencies of splenic epitope S510-specific V β 13-expressing CD8 T cell clonotypes fit a log-series distribution well (Fig. 2, Table II; in no case could a log series be rejected at $p = 0.05$). For each sample, α , a measure of CDR3 diversity, was calculated and used to estimate the abundance of TCR V β 13 cDNA clonotypes per 10,000 CD8 T cells and per spleen (Table II). These data can be extrapolated to the entire population of epitope S510-specific CD8 T cells, assuming that each V β subrepertoire is similarly diverse. Because $\sim 22\%$ of epitope S510-specific T cells express V β 13 (7), 1100–1500 different clonotypes per spleen must have recognized epitope S510. This calculation assumes that diversity is uniformly distributed among V β families. To address this directly, we performed similar analyses on V β 8-expressing, epitope S510-specific T cells using RNA harvested from mouse S510-1 (Tables I and II). This data set, considered by itself, differed just significantly from log series ($p = 0.043$), with the deviation being in the direction of too many rare clonotypes (that is, the clonotype frequency distribution is even more uneven than log series). However, the apparent significance of this deviation should not be overinterpreted, because we conducted eight simultaneous goodness-of-fit tests (Table II). The probability of at least one type I error (false positive) in eight tests

⁴ The on-line version of this article contains supplemental material.

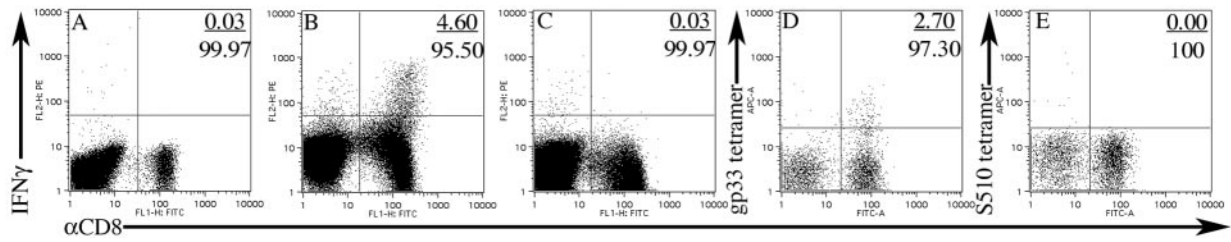


FIGURE 1. Specificity of sorting after staining with IFN- γ capture assays or MHC class I/peptide tetramers. A–C, Splenocytes were harvested from a naive mouse (A) or a mouse infected i.p. with JHM (B and C) and stained for IFN- γ production in response to peptide S510 (A and B) or no peptide (C) using an IFN- γ capture assay. D–E, Splenocytes were harvested from an LCMV-infected mouse and stained with tetramer gp33 (D) or irrelevant tetramer (JHM epitope S510 specific) (E). Gates for sorting were chosen to minimize nonspecific staining. The numbers show the percentage of CD8 T cells that are IFN- γ positive or negative (A–C) or tetramer positive or negative (D–E).

is not negligible (~33%), and if we had applied a sequential Bonferroni adjustment (20) to correct for the multiple tests, the deviation of the V β 8 data set from log series would not have reached significance (data not shown). Assuming that a log-series distribution is not inappropriate for these data, we calculated that there were ~206 V β 8-expressing CD8 T cell clonotypes per spleen. Because V β 8-expressing cells comprise 14% of epitope S510-specific CD8 T cells, this extrapolates to 1471 different epitope S510-specific clonotypes per spleen, in excellent agreement with the number calculated from the V β 13-specific population. By comparison, we previously calculated (7) that there were 350–500 different S510-specific clonotypes within the infected CNS. This difference arose in part because there are more epitope S510-specific CD8 T cells in the spleen than in the CNS, but the number of different cDNA clones per 10,000 epitope S510-specific T cells was also higher by 2- to 3-fold.

Diversity of splenic CD8 T cell responses to epitope S598

V β 5-expressing cells were analyzed because they made up ~28% of epitope S598-specific cells. Within this population, J β usage was diverse, with no J β 1 or J β 2 elements used preferentially (Supplementary Table II). Clonotype frequencies fit a log-series distribution well (Table II). As described above, we calculated that there were 160–250 V β 5-expressing epitope S598-specific CD8 T cell clonotypes per infected spleen, or 600–900 total S598-specific clonotypes per spleen (Table II). Six of 75 clonotypes were present in more than one mouse. We also attempted to analyze the epitope S598-specific T cell response in the CNS of mice with acute encephalitis using MHC class I tetramers. Despite poor cell recovery (see above), we were able to identify a limited number of CDR3 (~28 sequences from 5 brains). Three CDR3 were common to both splenic and CNS CD8 T cell samples. No CDR3 were com-

mon to the epitope S510- and S598-specific CD8 T cell populations.

Diversity of CD8 T cell response to LCMV epitope gp33

Blattman et al. (4) estimated that the precursor frequency to the LCMV-specific H-2D^b-restricted gp33 epitope was ~20 clonotypes per mouse, 1–2 orders of magnitude lower than we calculated for JHM-specific epitopes. Therefore, we determined the epitope gp33-specific precursor frequency using MHC class I (H2-D^b)/peptide tetramers to identify gp33-specific CD8 T cells in spleens of LCMV-infected mice (Fig. 1, D and E). Tetramers were used because epitope gp33 is presented by both H-2D^b and H-2K^b Ag (21) and the IFN- γ capture assay would detect both responses. We sequenced CDR3 from V β 8-expressing cells (comprising 30% of the H-2D^b-restricted tetramer gp33-positive cells (22)) and compared our results directly with those of Blattman et al. Clonotype frequencies fit a log-series distribution well (Table II). There were 1000–1200 different H-2D^b-restricted epitope gp33-specific T cell clonotypes in LCMV-infected mice (Table II; Supplementary Table III). This number is very similar to the number of clonotypes recognizing epitopes S510 or S598 in JHM-infected mice. The J β 1.1 element was used by ~50% of epitope gp33-specific T cell clonotypes, and aspartic acid was the first residue in a majority of the CDR3. In several cases, distinct nucleotide sequences encoded the same CDR3; this redundancy was not as common in the epitope S510- or S598-specific CDR3.

Testing appropriateness of log-series and uniform frequency distributions

Even after combining the results of goodness-of-fit tests across all data sets, we cannot reject the hypothesis of adequate fit to log-series distributions (Fisher’s method, $\chi^2 = 14.6$, df = 16, $p =$

Table I. Diversity and abundance of epitope-specific CD8 T cells in the spleen

Mouse	V β	No. Analyzed ^a	No. Species ^b	Abundance (no. copies) ^c						
				1	2	3	4	5–7	8–15	>16
JHM S510-1	V β 13	58	29	17	8	2	2	2	0	0
	V β 8	66	31	22	4	2	0	1	1	1
	V β 13	52	31	22	4	3	2	1	0	0
JHM S598-1	V β 5	49	29	16	7	1	1	2	0	0
	V β 5	51	29	19	7	0	0	1	1	0
	V β 5	56	23	15	2	1	2	1	2	0
LCMV gp33-1	V β 8	85	40	21	10	3	2	3	1	0
	V β 8	89	45	26	7	4	2	3	1	0

^a Number of cDNA clones analyzed per mouse.
^b Number of different cDNA clones identified.
^c Number of cDNA clones within each abundance class.

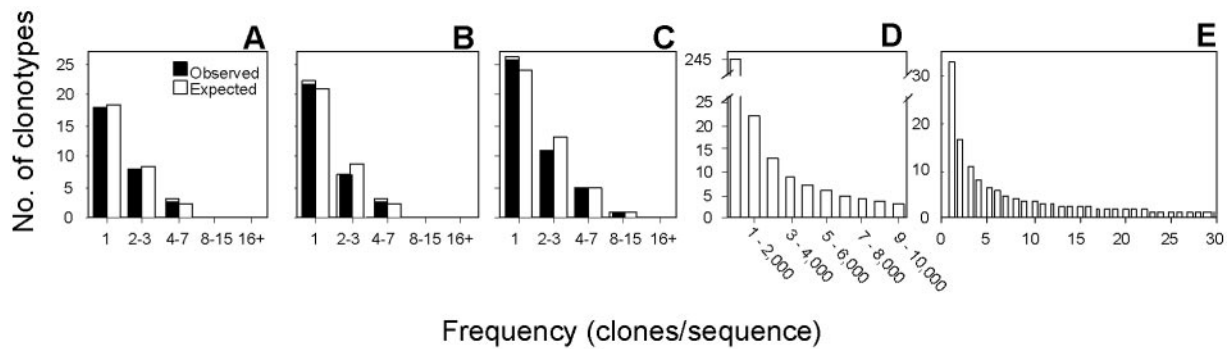


FIGURE 2. T cell clonotype diversities for representative mice. *A*, Frequency of T cell clonotypes (among our sequenced cells) recognizing epitope S510 in the spleen after i.p. infection with JHM; mouse S510-2. Expected frequencies are calculated using α estimated (Table II) from observed frequencies. *B*, Same for epitope S598, mouse S598-3. *C*, Same for epitope gp33, mouse gp33-2. *D* and *E*, Clonotype diversity for V β 8-expressing epitope gp33-specific T cells, mouse gp33-2, estimated for clonotype representation in the entire spleen. *D*, Broad frequency classes. *E*, Clonotypes inferred present at low copy number.

0.55). In contrast, we can strongly reject the hypothesis that clonotype frequencies follow a uniform distribution, as applied by other workers (4, 23). Across all data sets (Table II), clonotype frequencies differ strongly and significantly from expected under the uniform-distribution hypothesis (Fisher's method, $\chi^2 = 55.2$, $df = 16$, $p < 0.0001$).

Effect of sequencing additional cDNA clones on SE of α

In our analyses, we sequenced 50–100 cDNA clones and inferred the number of sequences from a fitted log series. Because this is a small fraction of the total number of epitope-specific CD8 T cells in the spleen, we estimated the likely improvement in our diversity estimates had we sequenced more clones, assuming that the underlying clonal diversity really is log series with α , as estimated. Benefits of further sequencing were modest: sequencing 1000 clones (vs 50) would decrease the SE of α only by a factor of 2 (Table III).

Discussion

Our analyses suggest that the number of CD8 T cell clonotypes in the naive mouse able to respond to a single epitope is substantially greater than previously reported. It is unlikely that our approach much overestimated precursor frequency. The IFN- γ capture assay detects cells that secrete IFN- γ in response to specific peptides, and not cells that secrete, instead, cytokines such as IL-4 or IL-10. Our assay depended on T cell activation, which could bias the results if transcription rates of TCR chain mRNA were very dif-

ferent among epitope-specific T cells; however, this is apparently not so (24, 25). Finally, our calculations do not consider redundancy in T cells expressing the same CDR3 RNA sequence, the contribution of α -chain usage to diversity, or the possibility that different sets of T cells might be activated after i.p. vs intranasal or intracranial inoculation. These caveats make it likely that we actually underestimated precursor frequencies.

It is unlikely that many of our sequences correspond to cellular or *Taq* polymerase-induced sequencing contaminants. First, gating was based on stringent criteria, with virtually no cells detected in the absence of specific peptide or peptide/tetramer, or in naive animals (Fig. 1). Second, the vast majority of epitope-specific sequences differed by more than 1 nt, except for the gp33-specific V β 8-J β 1.1 population. Because CDR3 were no more than 36 nt in length (Supplementary Tables I–III) and the *Taq* polymerase error rate was 0.12%, it is unlikely that these 1-nt differences in the gp33-specific V β 8-J β 1.1 population resulted from polymerase errors. It is more likely that these closely related sequences define a motif characteristic of this population and thereby serve as an indirect control for specificity.

Distributions of splenic epitope-specific CD8 T cells were well described by log series. Assuming that log-series distributions can be extrapolated to include clonotypes present at very low copy number, we predict that, for V β 8-expressing epitope gp33-specific CD8 T cells, 200 clonotypes will be very rare (<0.3% of the total epitope-specific cells per spleen), with 20–40 represented by single cells (illustrated in Fig. 2, *D* and *E*, for mouse gp33-2). Proving

Table II. Estimate of number of epitope-specific CD8 T cell clonotypes per infected spleen

Mouse	V β	V β 13, V β 5, or V β 8 Epitope-Specific Cells/Spleen (no.)	α (P) ^a	Frequency of V β 13, V β 5, or V β 8 clonotypes (per 10,000) ^b	Total No. of V β 13, V β 5, V β 8 Clonotypes/Infected Spleen (CI) ^b	Total No. of Clonotypes/Spleen ^c
JHM S510-1	V β 13	310,000	27.08 (0.60)	160	253 (170–333)	1,150
	V β 8	197,000	22.80 (0.043)	139	206 (139–271)	1,471
S510-2	V β 13	438,000	35.42 (0.73)	200	334 (226–437)	1,518
JHM S598-1	V β 5	160,000	24.70 (0.87)	148	217 (141–289)	775
	V β 5	350,000	26.27 (0.23)	156	249 (163–332)	889
	V β 5	1,030,000	14.59 (0.24)	95	163 (100–224)	582
LCMV gp33-1	V β 8	1,280,000	29.5 (0.95)	172	315 (223–404)	1,050
	V β 8	2,510,000	32.7 (0.78)	187	368 (264–469)	1,227

^a Values of p from a G test for goodness-of-fit to the estimated log-series distribution; $p < 0.05$ indicates significant departure of the data from the log-series distribution.

^b Numbers assume fit to the estimated log-series distributions. CI, 95% confidence interval.

^c Calculated assuming that cells expressing V β 13 or V β 8 (S510), V β 5 (S598), or V β 8 (gp33) comprise 22 or 14% (7), 28% (our unpublished data), or 30% (22) of epitope S510-, S-598-, or gp33-specific CD8 T cells, respectively.

Table III. Effect of sequencing more clones on confidence intervals for numbers of epitope S510-2-specific T cell clonotypes^a

Assumed No. Sequenced	Inferred Clonotypes/Spleen	Inferred Clonotypes/10,000 Cells
50	334 (138–439) ^b	200 (138–257)
100	334 (246–419)	200 (151–246)
200	334 (260–406)	200 (159–239)
300	334 (266–400)	200 (162–236)
400	334 (270–396)	200 (164–234)
500	334 (272–394)	200 (166–233)
1000	334 (279–388)	200 (169–230)

^a Based on recalculation of SE, assuming same estimate of α and sequencing a larger number of cDNA clones.

^b Numbers in parentheses are an approximate 95% confidence interval (estimate \pm 2 SE).

this directly would entail sequencing more cDNA clones than is feasible, so our estimates must be considered approximations to the number of different CDR3 actually present. However, it is not plausible that low-frequency clonotypes are entirely absent. Goodness-of-fit tests consistently rejected a uniform-distribution hypothesis for our clonotype diversity data, and this means that the likelihood of sampling clonotype frequency distributions like ours from T cell populations without substantial numbers of rare clonotypes was negligible. Of course, we cannot reject the possibility that the true T cell clonotype-frequency distribution is not truly log series, but merely log series-like with similar numbers of rare clonotypes; for this reason too, we emphasize that our estimates of CDR3 diversity are approximations.

The log-series distribution provides a good empirical fit for our data, but the biological mechanism for this distribution is not currently understood. Differences in the number of precursor cells for each clonotype in the naive population or in the duration of Ag exposure may contribute to the development of a log-series distribution. Although recent publications suggest that, once activated, T cells are committed to an extensive proliferative program, exceeding eight cellular divisions (26–28), differences in proliferative potential past eight divisions would result in a nonuniform distribution of T cell clonotypes. A nonuniform distribution would also result if T cell proliferation and differentiation were dependent upon the cumulative strength of the signal resulting from exposure to APCs and cytokines, as has also been postulated (29).

We detected limited commonality in CDR3 usage between different animals. Only 1 of 59, 6 of 72, and 6 of 59 of epitopes S510-, S598-, and gp33-specific T cell clonotypes (at the amino acid sequence level) were detected in more than one animal. These numbers agree with our previous study of epitope S510-specific CD8 T cells in CNS of mice with acute encephalitis (7) and with other studies showing T cell clonotypes recognizing individual epitopes to be largely nonoverlapping among animals (5, 30).

In previous reports, precursor frequencies to single epitopes were estimated at \sim 200 cells per spleen, or 20 different clonotypes (4–6), much lower than our estimates. We suspect that different assumptions about clonotype diversity distributions largely account for the discordant results. In one study, Bouso et al. (5) analyzed naive and immune repertoires of CD8 T cells responding to epitope A2–170 (H-2K^d restricted). However, their analysis differed from ours in that they assumed they had sequenced all clonotypes in the spleen (i.e., assumed a truncated frequency distribution with no unobserved clonotypes). Because approximately one-third of their clonotypes were observed only once, unobserved clonotypes were not unlikely. In addition, the HLA-A2-derived peptide differs from a homologous sequence in the H-2K^d Ag by only 3 aa, which may contribute to a limited CD8 T cell response. In the

second study, the precursor frequency of CD8 T cells responding to the LCMV gp33 epitope was estimated using transfer of transgenic T cells (4). However, the estimated clonotype diversity was depressed by the assumption that equal numbers of marker and endogenous T cells after exposure to Ag imply equal numbers of precursor cells. Because the marker T cells were monoclonal, that assumption will fail if the frequency distribution of endogenous T cell clonotypes is nonuniform. Our data (and see also Ref. 5) are clearly inconsistent with a uniform distribution. If the actual distribution were approximately log series (resembling our data and Ref. 5), the calculated number of precursors would be much higher and closer to our estimates (Table II).

Recently, Gorski and colleagues (31) showed that the distribution of memory human CD8 T cells recognizing an influenza A epitope included, like the distributions reported in this work, a large number of low-frequency clonotypes, which could be described by a power law-like distribution. When we fit these data to a log-series distribution, two of three data sets were well described, whereas for the third the number of clonotypes represented only once was somewhat underestimated. Although both methodologies describe the data, the log-series distribution is simpler and does not require that the data be separated into high- and low-frequency clonotypes before analysis.

Assuming that there are $2\text{--}5 \times 10^5$ different CD8 T cell clonotypes in the naive mouse spleen, our data suggest that only $2\text{--}5 \times 10^2$ different antigenic epitopes could be recognized, if each TCR clonotype recognizes a single epitope. This is much lower than previous estimates and suggests that T cells must exhibit extensive cross-reactivity for the immune system to respond to a large number of Ags (32). These data are also consistent with reports highlighting the existence of allogeneic CD8 T cell responses after viral challenge and of T cells generated in response to one virus, but able to recognize a second, heterologous virus (33).

Appendix

Sample calculation of log-series diversity

Imagine that 100 cells were sequenced, yielding a total of 10 distinct sequences, or clonotypes (3 clonotypes represented once; 1, twice; 1, 3 times; 1, 6 times; 1, 10 times; 1, 17 times; 1, 23 times; and 1, 36 times). The log-series fit uses only the total number of cells ($N = 100$) and of distinct sequences ($S = 10$). Iterative solution of the equation $S/N = [(1 - x)/x][-\ln(1 - x)]$ gives $x = 0.9730825$, and using $\alpha = N(1 - x)/x$, we find $\alpha = 2.7663$. We can test the adequacy of the log-series fit to these data by calculating expected frequencies, using form $s_i = \alpha x^i/i$: in this case, we find $s_1 = 2.69$, $s_2 = 1.31$, $s_3 = 0.85$, and so on. Notice that 73% of clonotypes are represented multiple times, and 27% are singletons. These expected frequencies (and also the observed frequencies) are then grouped into classes (1 time, 2–3 times, 4–7 times, 8–15 times, 16+ times) and goodness-of-fit assessed using standard G tests (in this example, the fit is excellent, with $G = 0.64$, $p > 0.85$). Finally, the expected number of clonotypes found in a larger sample of cells is given by $S' = \alpha \ln(1 + N'/\alpha)$: for example, if 1000 cells were sequenced, $S' = 16.3$ distinct clonotypes would be expected.

Acknowledgments

We thank Dr. V. Badovinac for splenocytes from LCMV-infected mice, Dr. K. Messingham for providing MHC class I/peptide gp33 tetramer, and Dr. J. Harty for critical review of the manuscript.

References

- Davis, M. M. 1990. T cell receptor gene diversity and selection. *Annu. Rev. Biochem.* 59:475.

2. Davis, M. M., and P. J. Bjorkman. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* 334:395.
3. Laouini, D., A. Casrouge, S. Dalle, F. Lemonnier, P. Kourilsky, and J. Kanellopoulos. 2000. V β T cell repertoire of CD8⁺ splenocytes selected on nonpolymorphic MHC class I molecules. *J. Immunol.* 165:6381.
4. Blattman, J. N., R. Antia, D. J. Sourdive, X. Wang, S. M. Kaech, K. Murali-Krishna, J. D. Altman, and R. Ahmed. 2002. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J. Exp. Med.* 195:657.
5. Bouso, P., A. Casrouge, J. D. Altman, M. Haury, J. Kanellopoulos, J. P. Abastado, and P. Kourilsky. 1998. Individual variations in the murine T cell response to a specific peptide reflect variability in naive repertoires. *Immunity* 9:169.
6. Maryanski, J., C. Jongeneel, P. Bucher, J.-L. Casanova, and P. Walker. 1996. Single-cell PCR analysis of TCR repertoires selected by antigen in vivo: a high magnitude CD8 response is comprised of very few clones. *Immunity* 4:47.
7. Pewe, L., S. B. Heard, C. C. Bergmann, M. O. Dailey, and S. Perlman. 1999. Selection of CTL escape mutants in mice infected with a neurotropic coronavirus: quantitative estimate of TCR diversity in the infected CNS. *J. Immunol.* 163:6106.
8. Stohlman, S. A., C. C. Bergmann, and S. Perlman. 1998. Mouse hepatitis virus. In *Persistent Viral Infections*. R. Ahmed and I. Chen, eds. John Wiley & Sons, New York, p. 537.
9. Bergmann, C. C., Q. Yao, M. Lin, and S. A. Stohlman. 1996. The JHM strain of mouse hepatitis virus induces a spike protein-specific D^b-restricted CTL response. *J. Gen. Virol.* 77:315.
10. Castro, R. F., and S. Perlman. 1995. CD8⁺ T cell epitopes within the surface glycoprotein of a neurotropic coronavirus and correlation with pathogenicity. *J. Virol.* 69:8127.
11. Perlman, S., R. Schelper, E. Bolger, and D. Ries. 1987. Late onset, symptomatic, demyelinating encephalomyelitis in mice infected with MHV-JHM in the presence of maternal antibody. *Microb. Pathog.* 2:185.
12. Wu, G. F., A. A. Dandekar, L. Pewe, and S. Perlman. 2000. CD4 and CD8 T cells have redundant but not identical roles in virus-induced demyelination. *J. Immunol.* 165:2278.
13. Pewe, L., G. Wu, E. M. Barnett, R. Castro, and S. Perlman. 1996. Cytotoxic T cell-resistant variants are selected in a virus-induced demyelinating disease. *Immunity* 5:253.
14. Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relationship between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12:42.
15. May, R. M. 1975. Patterns of species abundance and diversity. In *Ecology and Evolution of Communities*. M. L. Cody and J. M. Diamond, eds. Harvard University Press, Cambridge, p. 81.
16. Krebs, C. J. 1999. *Ecological Methodology*. Benjamin Cummings, Menlo Park.
17. Taylor, L. P., R. A. Kempton, and I. P. Woiwod. 1976. Diversity statistics and the log-series model. *J. Anim. Ecol.* 45:255.
18. Sokal, R. R., and F. J. Rohlf. 1981. *Biometry*. W. H. Freeman, New York.
19. Pewe, L., and S. Perlman. 1999. Immune response to the immunodominant epitope of mouse hepatitis virus is polyclonal, but functionally monospecific in C57BL/6 mice. *Virology* 255:106.
20. Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43:223.
21. Hudrisier, D., M. B. Oldstone, and J. E. Gairin. 1997. The signal sequence of lymphocytic choriomeningitis virus contains an immunodominant cytotoxic T cell epitope that is restricted by both H-2D(b) and H-2K(b) molecules. *Virology* 234:62.
22. Blattman, J. N., D. J. Sourdive, K. Murali-Krishna, R. Ahmed, and J. D. Altman. 2000. Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection. *J. Immunol.* 165:6081.
23. Casrouge, A., E. Beaudoin, S. Dalle, C. Pannetier, J. Kanellopoulos, and P. Kourilsky. 2000. Size estimate of the $\alpha\beta$ TCR repertoire of naive mouse splenocytes. *J. Immunol.* 164:5782.
24. Paillard, F., G. Sterkers, G. Bismuth, E. Gomard, and C. Vaquero. 1988. Lymphokine mRNA and T cell multireceptor mRNA of the Ig super gene family are reciprocally modulated during human T cell activation. *Eur. J. Immunol.* 18:1643.
25. Regnault, A., J. P. Levraud, A. Lim, A. Six, C. Moreau, A. Cumano, and P. Kourilsky. 1996. The expansion and selection of T cell receptor $\alpha\beta$ intestinal intraepithelial T cell clones. *Eur. J. Immunol.* 26:914.
26. Mercado, R., S. Vijn, S. E. Allen, K. Kerksiek, I. M. Pilip, and E. G. Pamer. 2000. Early programming of T cell populations responding to bacterial infection. *J. Immunol.* 165:6833.
27. Van Stipdonk, M. J., E. E. Lemmens, and S. P. Schoenberger. 2001. Naive CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat. Immun.* 2:423.
28. Kaech, S. M., and R. Ahmed. 2001. Memory CD8⁺ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. *Nat. Immun.* 2:415.
29. Lanzavecchia, A., and F. Sallusto. 2002. Progressive differentiation and selection of the fittest in the immune response. *Nat. Rev. Immunol.* 2:982.
30. Turner, S. J., G. Diaz, R. Cross, and P. C. Doherty. 2003. Analysis of clonotype distribution and persistence for an influenza virus-specific CD8⁺ T cell response. *Immunity* 18:549.
31. Naumov, Y., E. Naumova, K. Hogan, L. K. Selin, and J. Gorski. 2003. A fractal clonotype distribution in the CD8⁺ memory T cell repertoire could optimize potential for immune responses. *J. Immunol.* 170:3994.
32. Mason, D. 1998. A very high level of cross-reactivity is an essential feature of the T-cell receptor. *Immunol. Today* 19:395.
33. Welsh, R. M., and L. K. Selin. 2002. No one is naive: the significance of heterogeneous T-cell immunity. *Nat. Rev. Immunol.* 2:417.