Selective T Cell Expansion during Aging of CD8 Memory Repertoires to Influenza Revealed by Modeling

Yuri N. Naumov, Elena N. Naumova, Maryam B. Yassai and Jack Gorski

*J Immunol* published online 22 April 2011
http://www.jimmunol.org/content/early/2011/04/22/jimmunol.1100091

**Subscription**  
Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions**  
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**  
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Selective T Cell Expansion during Aging of CD8 Memory Repertoires to Influenza Revealed by Modeling

Yuri N. Naumov,*1 Elena N. Naumova,†1 Maryam B. Yassai,‡ and Jack Gorski‡

The aging of T cell memory is often considered in terms of senescence, a process viewed as decay and loss of memory T cells. How senescence would affect memory is a function of the initial structure of the memory repertoire and whether the clonotypes that make up the repertoire decay at random. We examine this issue using the T cell memory generated to the conserved influenza A epitope M158–66 which induces a strong, focused, but polyclonal CD8 T cell response in HLA-A2 individuals. We analyzed the CD8 T cell memory repertoires in eight healthy middle-aged and eight healthy older blood donors representing an average age difference of ~27 y. Although the repertoires show broadly similar clonotype distributions, the number of observable clonotypes decreases significantly. This decrease disproportionately affects low-frequency clonotypes. Rank frequency analysis shows the same two-component clonotype distribution described earlier for these repertoires. The first component includes lower frequency clonotypes for which distribution can be described by a power law. The slope of this first component is significantly steeper in the older cohort. Generating a representative repertoire for each healthy cohort allowed agent-based modeling of the aging process. Interestingly, simple senescence of middle-aged repertoires is insufficient to describe the older clonotype distribution. Rather, a selective clonotype expansion must be included to achieve the best fit. We propose that responses to periodic virus exposure may drive such expansion, ensuring that the remaining clonotypes are optimized for continued protection. The Journal of Immunology, 2011, 186: 000–000.
Materials and Methods

Subjects
Healthy middle-aged (43–49 y old) and older blood donors (68–84 y old) were identified as HLA-A2.1 by DNA typing. The blood samples were collected as part of a protocol approval by the Institutional Review Board of BloodCenter of Wisconsin.

M1 and M6-specific CD8 T cell cultures
Recall triplicate cultures were performed for 3 wk as described (13). Data from all three cultures was combined. It should be pointed out that in humans as in the mouse, simple recall cultures only work if the subject has been previously exposed or vaccinated to the epitope in question.

RNA isolation, cDNA synthesis, VB19 spectratyping, CDR3β plasmid subcloning, VB19-CDR3β plasmid isolation, and cycle sequencing
These were performed as described (10–13). From each recall analysis, ~400 sequences were generated (Table I). Only VB19 CD8 T cell clonotypes for which CDR3 was length 11 and encoded Arg98 and Ser99 at positions 5 and 6 (RS clonotypes) were used for statistical analysis.

Repertoire measures and characteristics
We analyzed the nature and number of responding T cells using the colony counting approach used previously (10, 12, 13). Clonotypes are defined by the unique V/D/J rearrangement generating the TCR β-chain, and the set of these clonotypes defines the repertoire. The total number of observed clonotypes is denoted as N. Each clonotype was observed in a number of colonies that vary from 1 to $R_{MAX}$, the set of observations due to the most frequently observed clonotype. Clonotypes observed only once are referred to as singletons, S. The total number of nucleotide sequences generated in an analysis is denoted as M. Two additional characteristics of the repertoire were calculated: the proportion of singletons, observations due to singletons over the total number of observations, $N_S/N$, and the proportion of observations due to the most frequent clonotype, $R_{MAX}/N$. These two characteristics correct for individual sample size and describe the extremes of the clonotype distribution.

A repertoire can also be viewed as an assembly of CD3 amino acid (CDR3aa) sequences involved in epitope recognition. In this case, the total number of distinct CDR3aa sequences observed is denoted as L.

Rank-frequency summary of a clonotype distribution
Because repertoires are defined as an assembly of clonotypes, there is the possibility of complex clonotype distributions, which were indeed observed in our research (12). As part of this analysis, the descriptive clonotype distributions are converted into quantitative clonotype ranks, $R_i$, where clonotypes of the same rank have identical number of observations within the cDNA libraries, and corresponding rank frequencies (12). For instance, clonotypes observed once (singletons) were assigned rank 1, clonotypes observed twice are rank 2, and so on. Maximal rank, $R_{MAX}$, is equivalent to the highest number of observations for the most frequent clonotype ($CDR3_{MAX}$), in which $C$ designates clonotype. To quantify rank-frequency summaries with rapid decay, we applied a power law equation, $y = a/x^b$, where $x$ is the rank and $y$ is the rank frequency. In the simplest situation, plotting a log/log transformation of the data, $\log y = \log a - b \log x$, should yield a straight line, in which the parameter $a$ indicates the frequency of observing single-copy clonotypes, and parameter $b$ describes the shape of the curve by indicating how rapidly the curve decays (19).

Our previous analysis indicated that the log-rank frequency summary took the form of a broken stick with one component being described as power law-like and a second component composed of a few clonotypes at different high ranks (12). To provide good fit to both parts of the rank-frequency summary, we expanded the model by allowing a separation of the curve at a critical point, $x_C$, as follows: $lny = lnlnx - bxlnx$, when $x < x_C$ and $lny = lnlnx - bx$, when $x \geq x_C$. Thus, the first portion includes the ranks represented by a high number of clonotypes (i.e., the extensive tail of singletons). Estimation of the parameters for each segment was performed by using a piecewise mixed-effects regression model (20). The estimate of $x_C$ was obtained iteratively by scanning the whole range of ranks. Because $x_C$ defines the critical rank, it is referred to as $R_C$ in the manuscript. Similarly, $b_1$ and $b_2$ are referred to as $B1$ and $B2$ in the text. The estimates are supplemented with 95% confidence intervals (CI) (insert in Fig. 3A) obtained from the statistical models. The quality of the model fit was assessed using standard methodology.

Diversity measures
We use a set of diversity measures (13) to examine diversity at the level of clonotypes and amino acid sequences in CDR3β (CDR3aa). A simple estimation of clonotypic frequency is a ratio of the total number of observations ($M$) and the total number of observed clonotypes ($N$) that yields the average number of observations per clonotype in a given sample ($V = M/N$). The measure of clonotype diversity we used in this analysis ($D_v$) is proportional to the number of clonotypes ($N$) with respect to a contribution of the most frequent clonotype to the whole repertoire: $D_v = R_{MAX}N/M - 1$.

Modeling of repertoire aging
Based on understanding critical phenomena in self-organized systems, we assume that consistently observed power-law distributions in memory repertoires are generated from a clonotypic birth-death process with repeated exposure to a pathogen (i.e., a multiplicative noise and a reinjection mechanism (page 313 in Ref. 21)) constrained by intermittent noise-amplifications in a globally contracting system (e.g., due to aging). In such a system, the evolution of a repertoire depends on the prior stage as a simple multiplicative recurrence process (page 307 in Ref. 21):

$$x_{t+1} = a(t)x(t),$$

where $x$ represents a clonotype abundance at time $t$, and $a(t)$ is a stochastic variable with probability $\Pi(a)$ that affects clonotype abundance. With no constraints, this process generates an ensemble of values $x(t)$ over all possible realizations of the multiplicative factors $a(0), a(1), a(2), \ldots, a(t)$, which is distributed as a log-normal distribution in the large $t$ limit as:

$$P(x) = \frac{1}{\sqrt{2\pi m}} \frac{1}{x^{2+e^t}}e^{-\ln(x)/m},$$

where $v = \ln(a) = \int_0^t da \ln a(t)$ and $D = \ln(a^2) - \ln(a^2) = \frac{1}{2Dt}$. Two constraints transform a log-normal distribution (2), also used by Kedzierska and coauthors (22), into a genuine power law distribution (page 308 in Ref. 21):

1) Global form of (1) must be contracting on average [i.e., $x(t) < x(t+1)$ for large $t$, equivalent to $v = 0$, the phenomenon we observe with aging.

2) While contracting, there are clonotypes that remain larger than a minimum value of $x_0 > 0$ and exhibit finite fluctuations preventing from contracting to zero.

The basic principles of repertoire modeling as a constrained birth-death process are described in Refs. 1 and 23. In brief, we define a system ($\Theta$, $\sigma$, $P$) that consists of a set of elementary events $\Theta$, $\sigma$-probability space, and probability measure $P$, according to Kolmogorov’s notations (23). Let $X$ be a set of unique clonotypes $\{x_1, x_2, \ldots, x_N\}$, where $N$ is a number of clonotypes and $x_i$ is a frequency of clonotype appearance. We assume that a set $\Theta$ consists of elementary transitions of a clonotype in a $\sigma$-probability space, measured by individual transitional probabilities $P$. In our experimental conditions, only the initial and the final stages of such transition are observable. To model such transitions in a discrete time, we assume that each clonotype at each stage possesses a unique characteristic probability to propagate and to be eliminated without replacement. The assumption of lack of replacement is based on our understanding of thymus involution. The majority of clonotypes are assumed to remain in the repertoire to retain immune memory.

Let $\theta_i$ be a propagation probability and $\omega_i$ be a removal probability for a clonotype $x_i$, so that $\theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$, and $\omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. The propagation probability relates to affinity/avidity or any other clonotype properties beneficial for immune response. The removal probability reflects a composite of probability of apoptosis and/or exhausted cell cycling. Using in silico experimentation, we assume that the process of repertoire development consists of a number of cycles, which can be a proxy to a calendar year, and a reinjection mechanism triggering intermittent amplifications. For simplicity, we assume that each cycle consists of two consecutive iterations: one for the contraction and one for the expansion. At each j-cycle, any clonotype can be removed with probability, $\omega_{\theta_i}$. We then allow each clonotype to contract at j-cycle with the probability, $\theta_{\theta_i}$ (e.g., a set will be extended by M copies of a clonotype $x_i$). To implement this process, for any given unique clonotype $x_i$,
at a $j$-cycle, we let $Y$ and $Z$ be random variables associated with a Bernoulli trial by defining it as follows: $Z$ (removed) = 0, $Z$ (kept) = 1. $Y$ (not propagated) = 0, $Y$ (propagated) = 1. Therefore, the simulated process consists of two draws: one for the repertoire contraction $z_{ij} \sim \text{Bernoulli}(\omega_{ij})$ and one for the repertoire expansion: $y_{ij} \sim \text{Bernoulli}(\theta_{ij})$. The Bernoulli sampling scheme has a straightforward interpretation; most importantly, it forms the basis for a birth-death process. For example, for the first cycle with $M = 1$, $y_{ijk}$ and $z_{ijk}$ denote $k$-event at $j = 1$ step for the $i$-clonotype (e.g., if $y_{ijk} = 0$ or $z_{ijk} = 0$, then $k = 0$; and if $y_{ijk} = 1$ or $z_{ijk} = 1$, then $k = 1$); thus, the propagation and removal processes can be presented in a following way:

$y_{100} = x_i$ is not propagated; only a single copy of $x_i$ at the cycle 1 is available; $y_{110} = x_i$ is propagated into two copies at the cycle 1; $y_{100} z_{i10}$: the only copy of $x_i$ is removed at cycle 1; $y_{110} z_{i11}$: two copies of $x_i$ are available for future propagation, and so on.

We considered a number of scenarios, in which we varied $M$ and implemented an agent-based modeling approach. We focused on two scenarios: with $M = 1$ to imitate a random contraction-expansion process (the null hypothesis), and with $M \approx x_i N$ to demonstrate the effect of selective expansion (the alternative hypothesis). The first scenario represents repertoire contraction by senescence with the removal of a constant number of sequences (cells) in each cycle and is based on current understanding of immunosenescence (18, 24, 25). To maintain an absolute number of T cells (homeostasis), the same number of remaining cells was randomly chosen to undergo duplication. This re-establishes the overall initial number of sequences, at which point the next cycle can begin. The second scenario replaces the equivalent expansion with an expansion based on the initial frequency of the clonotype chosen. We consider this a selective expansion based on clonotype frequency, which can also be considered a proxy for the avidity of the TCR–peptide–MHC interaction.

To illustrate the modeling procedure, we designed two repertoires imitating repertoires of an average middle-aged donor and an average elderly donor. The first repertoire consisted of 33 clonotypes with 329 sequences ($N_S = 9$, $R_{MAX} = 89$), and the second repertoire consisted of 22 clonotypes with 337 sequences ($N_S = 5$, $R_{MAX} = 138$). Together with the frequencies of clonotypes, these values represent constructed repertoires. We examined the conditions necessary to achieve the desired changes in the middle-aged constructed repertoire to arrive at the elderly constructed repertoire: a decrease in singletons synchronous with an increase in $R_{MAX}$ allowing for the same number of observations in a simulated repertoire. For each scenario, we performed 1000 simulations; in each simulation, we completed 50 iterations imitating an aging process. We performed sensitivity analysis and explored the trajectories and statistical properties of repertoire characteristics across simulated cycles. The simulated repertoires were further examined using rank-frequency summaries, as described above.

Results

Characteristics of the study cohorts

We analyzed eight middle-aged (average age 46.9 ± 1.96 y [mean ± SD]) and eight older blood donors (average age 74.3 ± 5.15). As regular blood donors, these are generally healthy individuals. None of the donors reported influenza-like symptoms in the month prior to donation or in the 2 y after donation.

Memory repertoires to influenza M158–66 in middle-aged and aging cohorts are polyclonal, yet the number of observed clonotypes is lower in older individuals

We performed recall cultures by stimulating PBMC with M158–66 for 3 successive weeks, with the cultures being restimulated with peptide-pulsed PBMC at the end of week 1 and 2. At the end of the third week, cells were collected from the cultures, cDNA prepared, and PCR amplified using VB19 and CB-specific primers, and the products were resolved on sequencing gels. This spectratype analysis showed the predominant presence of CD8 T cells with the expected CDR3 length of 150 bp. The PCR products were cloned into plasmid vectors and sequenced. We have described our approach to memory repertoire analysis using recall responses previously, including our method of counting clonotypes (10–13). It should be pointed out that simple in vitro recall responses in humans require previous exposure or vaccination to the recall Ag.

We define a responding repertoire as the collection of clonotypes (N) in the recall culture. A clonotype is identified by its CDR3 nucleotide sequences. The total number of observations made of each repertoire, $M$, is all of the sequences generated at the expected CDR3 length. The relative frequency of a particular clonotype is the ratio of the number of observations of that clonotype, $M_C$, and the total number of observations, $M$, ($M_C/M$). For high-frequency clonotypes, this methodology corresponds well with directly hybridizing the amplified cDNA with clonotype-specific probes (10, 26). Independent studies in which cultures were analyzed in parallel by the method described in this study and after tetramer selection showed that the high-frequency component of the repertoire is comparable (in preparation). We have previously defined repertoire measures and characteristics (13, 23). The average repertoire measures for each cohort are described in Table I. The average number of observations per subject in each cohort was close to the goal of 400. It can be seen that the average number of clonotypes in the older cohort is reduced. However, in this study, only clonotypes of CDR3 length of 11 aa encoding the canonical CDR3 motif of RS are considered (Table I). These clonotypes are recognized as being $M1_{58–66}$-specific, and the structure of a TCR–peptide–MHC complex is available (27, 28). They are the predominant species observed in MHC multimers analyses (11) and constitute approximately the same proportion of observations in each cohort (76.6% versus 79.1%). Although some of the non-RS clonotypes may be specific, the possibility that some of them may represent nonspecific noise leads us to focus on that part of the response that is well understood.

The recall repertoires for all 16 subjects are shown in Fig. 1. The RS-clonotypes are sorted in descending order of the number of observations, and the individual repertoires are sorted to generate a decreasing gradient for each cohort. Each repertoire consists of many clonotypes; a few clonotypes have a high number of observations, and many of the clonotypes are observed once. Overall, the patterns are similar between the two cohorts with high-, mid-, and low-frequency clonotypes in each individual. The fact that the gradient continues from the middle-aged to the older cohort already suggests that there might be fewer RS clonotypes identified in older individuals.

Using the repertoire measures described above, we define a number of repertoire characteristics. Two of the repertoire characteristics are normalized values for the measurements that represent the clonotypes observed once ($P_S = N_S/M$) and the most frequent clonotype, $P_{MAX} = R_{MAX}/M$. These two describe the contribution of the extremes of the clonotype distributions as observed in Fig. 1. Another characteristic represents a normalized assessment of clonotype abundance. The average number of observations per clonotype repertoire characteristics for each subject are shown in Fig. 2. Our experiments were designed to collect the same amount of CDR3 sequence data from each subject. However, there were some variations, and thus, the number of identified clonotypes ($N$) is normalized as the ratio of observation per clonotype for each subject ($V = M/N$). The values for the middle-aged cohort are clustered, whereas the older subject data are disperse (Fig. 2A). The average value of $V$ was significantly higher in older individuals compared with middle-aged donors. More observations per clonotype indicate a relative decrease (~40%) in the number of clonotypes in the circulating repertoire.
The proportion of the observations that was due to singleton clonotypes ($P_S$, Fig. 2B) was significantly higher in the middle-aged cohort, whereas the proportion of observations due to the most frequent clonotype ($P_{MAX}$, Fig. 2C) was significantly lower in the middle-aged cohort. This indicates that the two extremes of the repertoire are changing with age, in favor of less singleton contribution and more contribution from high-frequency clonotypes in the older cohort.

Clonotype diversity ($D_C$), a characteristic that incorporates richness ($N$) and unevenness ($R_{MAX}/M$), gave dispersion values for both cohorts (Fig. 2D), and there were no significant differences in the average value for each cohort. This repertoire characteristic is based on the assumption that a system is robust and resilient to distortions if it is rich in species (i.e., clonotypes) and their abundance is uneven (29), as is the case in this study. This lack of change of diversity is mirrored in the maintenance of the same general clonotype distribution observed in Fig. 1. The diversity does not change significantly because the decrease in clonotype richness in the elderly is offset by the increased unevenness of the repertoire.

Accompanying the decrease in the number of clonotypes is a decrease in the CDR3aa sequences present in the repertoire of older individuals. This is manifest as an increase in the number of observations per sequence, $V_{ML}$ (Fig. 2E). On average, the older cohort has ~66% of the number of CDR3 amino acid sequences in the repertoire.

These results indicate that the age-associated reduction in the observable number of M158–66 reactive clonotypes leads to a reduction in CDR3aa sequences involved in epitope recognition. Although the repertoires of the older donors have fewer clonotypes ($V$), a smaller contribution from singletons ($F_S$), a larger contribution from the most frequent clonotype ($PMAX$), and fewer available CDR3aa sequences ($V_{ML}$), the general shape of individual repertoires in both cohorts is still similar and consistent with what we consider a diverse repertoire.

**Rank-frequency analysis**

The analysis in the previous section was based on a simple description of repertoire characteristics. A more detailed description can be obtained in the form of individual rank-frequency summaries. For a rank frequency analysis, we assign each clonotype to a countable entity, the clonotype rank. Clonotype ranks pool all of the clonotypes present at a particular frequency of observation and range from clonotypes observed once to the largest number of observations for a clonotype, $R_{MAX}$. We have previously shown that for most individuals, a log-log plot of the rank versus rank frequency has a component that starts with the highest frequency for rank 1 and decreases linearly until a critical point, after which the remaining high ranks are represented by fewer clonotypes (12). We performed such an analysis and then estimated the average trajectories for each cohort by fitting the data by a piecewise mixed-effects model. Each trajectory (curve) thus represents the average repertoire shape of each age cohort.

The average trajectories for rank-frequency summaries for each cohort are shown in Fig. 3A. As can be seen, there are two components to the clonotype distribution, both of which have their individual slopes (parameters $B_1$ and $B_2$) and are divided by the critical rank $RC$ (dotted vertical line in Fig. 3A). The curve to the left of $RC$ represents the large number of clonotypes at low ranks, starting with singletons, and shows a steady decline in occurrence of clonotypes as rank increases. The rate of decrease $B_1$ differs significantly between the two cohorts: 0.42 (95% CI: 0.37, 0.47) for the middle-aged cohort versus 0.25 (95% CI: 0.19, 0.34) for the older cohort.

### Table I. Clonotype characteristics for middle-aged and older cohorts

<table>
<thead>
<tr>
<th>Repertoire characteristics</th>
<th>Middle-Aged Cohort</th>
<th>Older Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of all unique clonotypes, $N_{ALL}$</td>
<td>64.25 ± 17.56*</td>
<td>34.38 ± 8.12</td>
</tr>
<tr>
<td>No. of observations of all unique clonotypes, $M_{ALL}$</td>
<td>431.00 ± 129.87</td>
<td>412.38 ± 61.46</td>
</tr>
<tr>
<td>RS-clonotype repertoire characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of unique RS clonotypes, $N$</td>
<td>38.25 ± 12.60*</td>
<td>22.63 ± 8.45</td>
</tr>
<tr>
<td>No. of observations of RS clonotypes, $M$</td>
<td>329.75 ± 97.11</td>
<td>326.13 ± 128.35</td>
</tr>
<tr>
<td>Proportion of RS clonotypes to all clonotypes, $N/N_{ALL}$</td>
<td>0.60 ± 0.14</td>
<td>0.66 ± 0.17</td>
</tr>
<tr>
<td>Proportion of RS observations to all observations, $M/M_{ALL}$</td>
<td>0.78 ± 0.11</td>
<td>0.78 ± 0.26</td>
</tr>
<tr>
<td>No. of singletons, $N_S$</td>
<td>11.88 ± 5.64</td>
<td>7.75 ± 3.41</td>
</tr>
<tr>
<td>Maximum number of observations per a unique clonotype, $R_{MAX}$</td>
<td>64.00 ± 22.03*</td>
<td>120.88 ± 70.44</td>
</tr>
<tr>
<td>No. of CDR3β amino acid sequences, $L$</td>
<td>18.13 ± 4.42*</td>
<td>11.38 ± 4.17</td>
</tr>
<tr>
<td>No. of observations of the most frequent CDR3β amino acid sequence, $R_{MAX,αa}$</td>
<td>110.25 ± 35.05</td>
<td>156.00 ± 86.67</td>
</tr>
</tbody>
</table>

Values are mean ± SD.

*p < 0.01.
older cohort, \((p < 0.01)\) (inset in Fig. 3A). The curve to the right of \(R_C^{C}\) represents fewer clonotypes present at high ranks (i.e., clonotypes observed very frequently). The rate of decrease given by the parameter \(B_2\) is similar between the two cohorts: 0.92 (95% CI: 0.79, 1.06) for middle-aged versus 0.85 (95% CI: 0.79, 0.91) for the older cohort \((p > 0.05)\). The age-related decrease in the singleton clonotypes and the increase in \(R_{MAX}\) already described in Fig. 2 can be observed in this representation. The rank-frequency analysis shows that rate of decrease is lower in the first component (B1) of the older cohort, which indicates that the first power law component takes a smaller proportion of the entire repertoire in this cohort. As a result, there is a change in the critical rank that divides the two components.

Requirement for frequency-based selection in models of repertoire aging

Our data describe the repertoire in two age-based cohorts. However, there is an implied dynamic process that should lead from one

FIGURE 2. Distributional characteristics of the recall repertoire for the two aged cohorts. The mean and SD of the repertoire characteristics for each cohort are given for each characteristic. The \(p\) values are shown for difference of the mean values of the two cohorts using the nonparametric \(t\) test. A, Average number of observations per clonotype \((V = M/N)\), \(B\), Proportion of singletons; the number of observations of singletons divided by the total number of observations \((P_S = N_S/M)\). \(C\), Proportion of observations contributed by the highest frequency clonotype \((P_{MAX} = R_{MAX}/M)\). \(D\), Clonotype diversity \((D_C = R_{MAX}N/M - 1)\). \(E\), Average number of observations per CDR3aa sequence \((V_{ML} = M/L)\) (RS-clonotypes with different flanking amino acids).

FIGURE 3. Clonotype rank-frequency analysis of the memory repertoires. A, Predicted clonotype rank-frequency trajectories for the average middle-aged (red) and older (blue) cohort plotted against the rank. The critical rank, \(R_C^{C}\), divides the two portions of the trajectory and is denoted by the dotted vertical lines. The \(\ln(R_{MAX})\) defines the average maximum number of observations for a clonotype on a log-scale. The parameters B1 and B2 define the change in rank in the first and second portion of the curve, respectively. B, Aging without selection. As described in the Materials and Methods, the best fit to the data is shown for a process in which the average middle-aged repertoire undergoes random contraction and re-expansion. The lines shown are for 15 (orange), 25 (violet), and 35 (green) y. C, Aging with selection. The best fit is shown for a process in which sequences to the right of \(R_C^{C}\) are given an extra division during the expansion process. Clonotype ranks (log scale) are shown on the \(x\)-axis. The cumulative rank frequencies within repertoires are shown on the \(y\)-axis.
cohort to the other. We modeled this process by using repertoire parameters from the experimental data to generate discrete repertoires in silico for the middle-aged (Fig. 3B, open circles) and older (closed circles) cohorts. These in silico repertoires were very similar to the generalized repertoire calculated from our data (Fig. 3A).

We then used an agent-based approach to ask if we could reconstruct the older cohort pattern from the middle-aged pattern by implementing an iterative contraction process in which T cells are removed at each cycle (e.g., in 1 y) and then renormalized by expansion of the remaining cells. The removal would be equivalent to senescence and the renormalization equivalent to non-Ag–driven homeostatic expansion. The peripheral expansion is necessary to explain the normal lymphocyte blood count and CD8 T cell proportion of the older cohort (data not shown). The maintenance of M1.<ampmath>58.66</ampmath>-specific cell numbers in PBMC of older populations has been reported (16).

A number of contraction and expansion parameters were analyzed and the best fit achieved by eliminating four randomly selected sequences (T cells) during the contraction phase and then selecting four random sequences for expansion. The model ignores the addition of new clonotypes as the thymus in both these age cohorts has involuted and is producing very few new T cells if any (30, 31). Using these parameters for simulation, the model repertoire was obtained by averaging 500 simulated trajectories. Repertoire data are shown for three time periods (Fig. 3) corresponding to a cycle times of 15, 25, and 35 y (orange, violet, and green lines, respectively). The series of curves have similar shapes showing a decrease in the frequency of singleton (<ampmath>N_S</ampmath>) clonotypes and a change in slope B1. The 25-y curve (violet line) has changed <ampmath>N_S</ampmath> to the same extent as the observed value. However, other aspects of the clonotypic distribution achieved after this time do not correspond to the constructed values of the older cohort. The loss of clonotypes, the B1 parameter, and the corresponding critical point have not changed to the extent observed in the actual data. The <ampmath>R_{MAX}</ampmath> is lower than the observed data.

By changing the rate of senescence (and re-expansion), one can match the model at the 25-y time point to one of the other aspects of the constructed older data (e.g., extent of loss of clonotypes), but under these conditions, all of the other aspects do not match. This indicates that a simple contraction process in which random T cells are removed and the remaining T cells re-expand nonselectively cannot explain the actual observations.

One alternate that was modeled with interesting outcomes was the effect of combining a selected expansion with senescence. The process is the same until the expansion, at which point the randomly chosen cells expand on the basis of their initial rank. If a sequence is picked that ranks higher than <ampmath>R_C</ampmath>, an additional copy is added to the repertoire, and there is one fewer pick. With this simple model, the 25th cycle results in a repertoire that on average matches the number of singletons, the slope B1, <ampmath>R_C</ampmath>, and clonotype loss that was observed for the average trajectory of the older cohort (Fig. 3C, violet line). Although a good fit was achieved for a point estimate of <ampmath>R_{MAX}</ampmath>, the predicted value had a relatively wide CI. Thus, after 25 y in silico, senescence coupled with selective expansion leads to a repertoire for which the number of clonotypes and rank frequency distribution are very similar to those observed for the older cohort.

**Discussion**

The effectiveness of antiviral immunity results from well-organized innate and adaptive immune responses that control viral replication and eventual viral clearance. Inherent in the adaptive responses are the clonal expansion and contraction of B and T cells specific to the particular virus. Age-associated reduction of thymic function leads to greatly diminished naive T cells and a greater reliance on pre-existing memory. Aging after thymic involution can be considered as a senescence process by which the number and function of memory T cells decreases with time. This immunosenescence has been proposed as a key factor in the high susceptibility of the elderly to infectious diseases and cancer (32).

Understanding the process of immune aging is important but hampered by the polymorphism of the subjects and their different immune histories. Longitudinal studies provide an inbuilt control, but longitudinal studies covering 20–30 y periods are very difficult to conduct or fund. Therefore, modeling the aging process will have to play an important role in this regard. Modeling of changes in the immune system falls into two major categories: models describing the dynamics via differential equations (33, 34) or via various simulation experiments (23) including agent-based models (35). The former aims to carefully describe the behaviors of cell populations over time and the specific interplay of their dynamics. The latter approach formulates a series of assumptions and allows simulating the process in discrete steps.

In this study, we used agent-based modeling to age a generalized middle-aged repertoire to examine what parameters would provide the best fit with the generalized older repertoire we derived from our cohort analyses. We then imitate the process of aging, which is unrealistic to track in real time, in a well-controlled experiment in silico with proper attributes of sensitivity analysis, uncertainty considerations, and quality control. The recall CD8 T cell repertoire to influenza A provides an excellent opportunity to measure and model aging because it is polyclonal and the clonotype distribution is complex. This provides a number of characteristics that can be compared and contrasted between cohorts and incorporated in the model. From our comparison of simple repertoire characteristics, we measured an ∼2-fold decrease in the number of observable responding clonotypes in the older cohort, as might be expected from the aging process.

This observation is in keeping with our longitudinal analyses of middle-aged individuals, which revealed a decrease in the number of RS clonotypes over a 7–10 y period (13). However, it is also possible that the decrease in observable RS clonotypes is due to a loss of their proliferative capacity in vitro. If the in vitro observation represents the in vivo status of these cells, then the immunosenescence modeled in this study would not represent a loss of clonotypes, but rather of their ability to proliferate. It is most likely that the measured differences are a result of both processes.

There was an increase in the rank of the most frequent clonotype with age and a decrease in the proportion of singletons. The only characteristic that did not change was the calculated clonotype diversity, <ampmath>D_C</ampmath>, which incorporates abundance and unevenness. In the ecological literature, both of these characteristics are considered important when measuring diversity and have been used in measuring T cell repertoire diversity (13, 36). The equivalent diversity measure indicated that the repertoires still were polyclonal and still uneven.

To further investigate the unevenness of the repertoire at a more detailed level, rank-frequency analysis was performed. This revealed additional differences, as the slope of the power law component decreased in the older cohort, leading to a change in the critical rank that divides the two components of the clonotype distribution. If one considers the broken stick description of the middle-aged repertoire (12) and observes that there is a change between the two components, a number of possible scenarios present themselves. The proportion of singletons may remain the same, and only the higher frequency clonotypes in the first component may be promoted to the second component. In such a
scenario, the first point in the log-log plot would remain the same, the slope would decrease, and the critical point would shift leftward. If, in contrast, the singletons were more susceptible to loss, then the first point would decrease, the slope would change, but the critical point would remain the same. Finally, if the power law component were to senesce, the slope would remain the same, but a change in the first point and a shift in the critical point would be observed. None of these three scenarios describe what we observed in the older cohort. As the number of possibilities could be very high, we performed modeling to explore potential mechanisms to transition from the middle-aged to the elderly distributions.

The in silico simulation indicated that simple senescence was insufficient to explain the difference between the two generalized cohort descriptions (Fig. 3B). This is to some extent surprising as the simple senescence model might be expected to affect the entire first power law component. What this model shows is that by eliminating cells from a power law distribution, the singleton component is more sensitive. This can be simply stated as once a clonotype is gone, it cannot come back. Eliminating one copy of a clonotype in the central or high frequency portion will not greatly change the overall distribution, as one clonotype loses one copy and other gains a copy. In the context of higher frequencies, this does not represent a major perturbation.

An alternative model explored the assumption that the probability of expansion is related to frequency (Fig. 3C), with interesting results. The in silico aging for 25 yr resulted in a clonotype distribution that closely resembled the generalized older cohort. The assumption of a relation between rank and cell proliferation would be expected if the selective expansion was actually due to an exposure to the M158–66 epitope, and such an exposure preferentially mobilizes the higher frequency clonotypes, both of which are reasonable assumptions. An alternative is that there is preferential homeostatic expansion of clonotypes that have higher avidity for M158–66. We consider this possibility less likely, as the mechanism for such a correlation does not appear straightforward. We therefore propose that exposure to influenza during aging has a major effect on the memory repertoire, changing the middle-aged distribution of clonotypes to resemble those observed in the older cohort. On the basis of this model, we predict that experiments that measure repertoires before and after exposure should reveal these changes.

This relatively simple model shows the importance of a selective expansion step but could be refined further. For a simple fractal structure, it could be easily shown that a reduction in singletons can be compensated by an increase in dominant clonotypes in the following way: $\ln R_{MAX}(t + 1) = \frac{\alpha}{\gamma} \ln R_{MAX}(t)$ and $\ln N_S(t + 1) = \frac{\beta}{\gamma} \ln N_S(t)$, so if $\gamma = 1$ then $\alpha = \frac{1}{\beta}$ where $\alpha$ and $\beta$ reflect rates of change in log-abundance of most dominant clonotypes and singletons, corrected for $\gamma$. which is a change in a repertoire size, $M$. Similarly, the conditions that predict the dynamics of a repertoire with a multifractal structure can be shown. For a two-component distribution with $M_L$ and $M_H$ representing the number of sequences for low-frequency and high-frequency clonotypes, respectively, when $M(t + 1) = M(t)$, we observe a compensatory mechanism described by the equation:

$$M_L(t + 1) + M_H(t + 1) = \alpha \beta C \ln R_C(t) \ln N_C(t) + \alpha C \beta \ln N_R(t) \ln R_{MAX}(t),$$

where $\alpha$ and $\beta$ reflect rates of change in log-abundance of the most dominant clonotypes and singletons and where $\alpha_C$ and $\beta_C$ control the shift of $R_C$. Because Eq. 4 includes all of the variables and parameters that affect the two-component description of the clonotype distribution, models based on this equation should be able to describe all possible transformations of one repertoire into another.

Eq. 4 describing the relationship between the two compartments has important implications in considering diversity. Our diversity measure $D_C$ is sensitive to expected changes in both compartments, whereas most other measures are primarily aiming to distinguish an observed distribution from a more even or uniform shape (36).

This model could be further expanded to cover longer time intervals, inclusion of a proportion of naive responses similar to ideas expressed in Ref. 35, include frequency and intensity of pathogen encounters, and incorporate the life span of T cell clones (33). Although such additions would have a high demand for computational power in an agent-based model framework, they are worth pursuing. Another refinement would be to further examine the relation between clonotype frequency and the selective expansion in a substantially large sample by paying a special attention to changes in the low-frequency compartment.

Direct studies of clonotypical repertoires of Ag-specific T cells over 20–30 y have yet to be performed. Previous modeling studies involved demonstrating the changes in immune memory with aging and potential acceleration in aging associated with chronic infection (34, 37). Our data and modeling at the clonotype level can provide a benchmark for studies that use other older cohorts, such as even older healthy individuals, or individuals admitted to hospital. If maintaining a two-component repertoire distribution does provide enhanced protection, this should become evident from such studies. Our data can also be used for comparison of repertoire aging studies in individuals with other HLA types.

The changes we have modeled are associated with two healthy populations and could be indicative of the clonotypic memory T cell repertoire structures associated with continued health. None of the individuals in this study reported influenza-like illness during a period of 2 y after these samples were obtained. It will be important to extend these studies of how T cell memory ages at near-extreme (i.e., frequent acute infections) and extreme (i.e., frequent acute infections coupled with comorbidity) conditions. If our data reflect an optimum health pattern, then boosting cytotoxic T cell responses early in life to generate complex memory repertoires may be very valuable, something that is not the target of current vaccination practice.

Acknowledgments
We thank Erica Reed, Melissa Unruh, Wendy Demos, Dmitry Bosenko, Kalyani Kota, and Pavan Katikaneni for excellent technical assistance.

Disclosures
The authors have no financial conflicts of interest.

References