The somatically generated T cell receptor CDR3α contributes to the MHC allele specificity of the T cell receptor

Philippa Marrack,[1,2,3] Sai Harsha Krovi,[3] Daniel Silberman,[2,3] Janice White,[2] Eleanora Kushnir,[2] Maki Nakayama,[3,6] James Crook,[4] Thomas Danhorn,[4] Sonia Leach,[2,4] Randy Anselment,[4] James Scott-Browne,[5] Laurent Gapin,[3] and John Kappler,[1,2,3]

[1]Howard Hughes Medical Institute, Denver, CO 80206;

[2]Department of Biomedical Research, National Jewish Health, Denver, CO 80206;

[3]Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO 80045;

[4]Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, Colorado 80206

[5]La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037;

[6]Barbara Davis Center for Childhood Diabetes, University of Colorado School of Medicine, Aurora, CO 80045.

ABSTRACT

Mature T cells bearing $\alpha\beta$ T cell receptors react with foreign antigens bound to alleles of major histocompatibility complex proteins (MHC) that they were exposed to during their development in the thymus, a phenomenon known as positive selection. The structural basis for positive selection has long been debated. Here, using mice expressing one of two different T cell receptor β chains and various MHC alleles, we show that positive selection-induced MHC bias of T cell receptors is affected both by the germline encoded elements of the T cell receptor α and β chain and, surprisingly, dramatically affected by the non germ line encoded CDR3 of the T cell receptor α chain. Thus, in addition to determining specificity for antigen, the non germline encoded elements of T cell receptors may help the proteins cope with the extremely polymorphic nature of major histocompatibility complex products within the species.

INTRODUCTION

Many T lymphocytes in the body express clonally distributed T cell antigen receptors composed of alpha and beta chains (TCRs) that react with peptides derived from pathogens bound in a groove on the surface of host major histocompatibility proteins (MHCs). The genes encoding these MHC proteins are the most polymorphic genes in a given species. Most of the polymorphisms tend to be concentrated within the residues that line the peptide-binding groove of the molecules. Hence, in general, different MHC alleles within a species preferentially bind, and present to TCRs, different peptides from any given invading organism. Thus the pathogen is unlikely to mutate such that none of its peptides bind to any of the MHC proteins expressed within the target species and the immune responses of at least some individuals within the infected species will thus be able to deal with the invading pathogen.

Many years ago another consequence of MHC polymorphisms was recognized. The allelic variants of MHC expressed in one individual are very frequently recognized by 1% or more of the T cells of other individuals expressing different MHC alleles, a phenomenon called "alloreactivity". While differences in bound peptides play an important role in alloreactivity(Hunt et al., 1990, Crumpacker et al., 1992), structural studies show that some of the allelic variations in MHC proteins themselves interact with the TCRs of alloreactive T cells (Grandea and Bevan, 1993, Archbold et al., 2008, Colf et al., 2007).

Experiments have shown that T cells in one individual are more likely to react with foreign peptides bound to the grooves of self MHC than to foreign peptides bound to foreign MHC (Fink and Bevan, 1978, Zinkernagel et al., 1978, Kappler and Marrack, 1978, Sprent, 1978). This phenomenon, known as positive selection, is caused by the fact that thymocytes are allowed to develop into mature T cells only if the TCR they bear reacts with low affinity/avidity with MHC proteins bound to self peptides in the thymus (Sprent et al., 1988, Ashton-Rickardt et al., 1994,

Sebzda et al., 1994, Hogquist et al., 1994). Paradoxically, in an apoptotic process termed "negative selection", the thymus generally weeds out T cell progenitors that react with too high affinity/avidity with self MHC plus self peptide, thus preventing the maturation of many potentially self reactive T cells (Kappler et al., 1987, von Boehmer et al., 1989). Thus the collection of TCRs on mature T cells in any individual bears the footprint of positive selection, reacting almost undetectably with self MHC bound to self peptide and being more likely to react with foreign peptides bound to alleles of MHC to which they were exposed in the thymus than to peptides bound to unfamiliar MHC (Fink and Bevan, 1978, Zinkernagel et al., 1978, Kappler and Marrack, 1978, Sprent, 1978, Hunig and Bevan, 1981).

Mutational and structural studies have shown that the alpha and beta chains that comprise TCRs each usually engage MHC + peptide via three complementary determining loops (CDRs 1,2 and 3) (Garcia et al., 1996, Reinherz et al., 1999, Dai et al., 2008). For both the TCR alpha and beta chains, two of these loops, CDR1 and CDR2, are encoded by the germ line TRAV (for the TCRα chain) and TRBV (for the TCRβ chain) genes. The third, CDR3, loop for each chain, on the other hand, is produced during TCR gene rearrangement as the cells develop in the thymus (Davis, 1985). Thus, the sequence coding for CDR3α ,for example, is created when one of many TRAV gene segments rearranges to fuse with one of the many TRAJ gene segments with the total number of possible CDR3α sequences increased by removal and/or addition of bases at the joining points of TRAV and TRAJ (Gellert, 2002, Cabaniols et al., 2001, Moshous et al., 2001, Lu et al., 2008). This process creates the DNA coding for the entire Vα domain. The stretch of DNA coding for CDR3β is constructed along the same lines, by joining of one of a number of TRBV, TRBD and TRBJ gene segments, again with bases removed or introduced at the joining points to form the CDR3 loop of the complete Vβ domain.

T cell receptors react with a surface of MHC plus peptide that is made up of amino acids from two MHC alpha helices that flank the groove where the peptide is bound. In the dozens of solved structures of TCRs bound to peptide-MHC complexes a pattern has emerged in which, usually, the germ line encoded TRAV and TRBV CDR1 and CDR2 loops focus on the MHC alpha helices whereas the somatically generated TCR CDR3 loops focus on the peptide (Garcia et al., 1999, Reinherz et al., 1999, Colf et al., 2007, Rudolph et al., 2006). Thus one might predict that positive selection selects TCRs that react well with peptides bound to self rather than foreign MHC by picking out TCRs bearing TRAVs and TRBVs that react favorably with self MHC. Indeed there is evidence that this is the case (Pircher et al., 1992, Merkenschlager et al., 1994, Sim et al., 1996). However, different MHC alleles will present different self peptides to developing thymocytes, therefore it is also possible that it is the presented peptide rather than the MHC protein itself, that governs the allele bias of positive selection. Again some evidence suggest that the selecting peptide is crucial (Ignatowicz et al., 1996, Tourne et al., 1997, Nikolic-Zugic and Bevan, 1990, Hogquist et al., 1994, Ashton-Rickardt et al., 1994, Wong and Rudensky, 1996, Barton et al., 2002), results that favor the idea that CDR3 sequences dominate positive selection. Understanding of this issue is complicated by the cooperative nature of TCR interactions with its ligands, by which an interaction at one site on the TCR/MHC/peptide surface adjusts interactions elsewhere (Mazza et al., 2007, Baker et al., 2012, Adams et al., 2016) and a study that indicated that the entire sequence of the TCRα chain, including the TRAV, TRAJ and CDR3α, is involved in positive selection (Merkenschlager et al., 1994).

We set out to resolve these issues. We analyzed the TCRα repertoires of naïve CD4 T cells in mice that each expressed one of two TCRβ chains, DOβWT or DOβ48A (Scott-Browne et al., 2009), and a single MHCII protein, IA, of alleles b, f or s (for simplicity and ease of reading we will use IA to describe what are often termed I-A proteins, and we will not use superscripts to

denote MHC and IA alleles). As predicted by previous studies (Pircher et al., 1992, Merkenschlager et al., 1994, Sim et al., 1996), the frequency with which mature T cells used different TRAVs was indeed affected to some extent by the MHCII allele on which they were positively selected and by the coexpressed TCRβ. Likewise the TRAJs used were affected by the selecting MHCII allele and coexpressed TCRβ, but demonstrated unexpected biases towards use of the TRAJs that were furthest from the TRAV locus.

Most surprisingly, however, the CDR3α sequences differed markedly depending on the MHCII allele and partner TCRβ in the mouse. This was true even if we compared, between MHC alleles, the TCRα sequences constructed from rearrangements involving the same TRAVs and TRAJs, indicating that the non germ line encoded portions of CDR3α are involved in MHCII allele specific selection.

RESULTS

The generation and properties of mice expressing a single TCR beta chain

The impact of positive selection on the TCR repertoire of mature T cells cannot be understood by sequencing only the expressed TCRα or TCRβ chains. This is because others and we have found a fairly high percentage of individual TCRα or TCRβ sequences are expressed in animals regardless of their MHC haplotype (Robins et al., 2010, Warren et al., 2011, Liu et al., 2014) **(Supplementary Table 1)**. Presumably this is at least in part possible because each individual chain is paired with a different partner(s) in animals with different MHC alleles. Therefore, the pairs of TCR chains expressed in individual T cells must be known in order to understand the impact of thymus selection on the TCR repertoire of mature T cells.

The T cells in any given mouse or human have been reported to bear, collectively, more than 10^5 different TCRα and about the same number of different TCRβ chain sequences (Venturi et al., 2011, Li et al., 2016). Thus the T cells might bear up to 10^{10} different combinations of these chains. Although methods for sequencing and accurately pairing the TCRα and TCRβ (or immunoglobulin heavy and light chain) RNAs from many individual T (or B) cells have been described (Tan et al., 2014, DeKosky et al., 2013), in our experience (Munson et al., 2016) these are still not able to cope with the large numbers of individual chains and combinations we expect in normal animals. Therefore we decided to limit our analyses to the T cells in animals that expressed a single TCRβ and any possible TCRα. This choice has two advantages. It allowed accurate knowledge of the TCRβ on the T cells and, because it was expected that only a limited number of TCRα chains can be positively selected with a single TCRβ, it limited the numbers of different TCRα sequences we expected to find in the mice (Merkenschlager et al., 1994, Fukui et al., 1998, Hsieh et al., 2006).

We chose two TCRβ chains for these experiments **(Supplementary Table 2)**. These were the TCRβ originally isolated from a T cell hybridoma constructed from BALB/c T cells specific for IA^d or IA^b bound to a peptide from chicken ovalbumin, the DOβWT TCRβ (White et al., 1983) and the same TCRβ with a mutation in its TRBV region such that the tyrosine at position 48 was changed to an alanine, DOβ48A (Scott-Browne et al., 2009). This mutation reduces the ability of the TCRβ chain to react with the alpha chain alpha helix of MHCII and with the alpha1 alpha helix of MHCI. The chain was chosen for our analyses because we thought that the TCRα sequences that could successfully overcome the deficits in MHC recognition by the TCRβ chain might more clearly illustrate the properties of the TCRα needed for successful positive selection.

The goal of these studies was to find out how the allele of MHC involved in thymic selection affects the sequences of the TCRs on the selected T cells. To achieve this we studied TCRs on naïve CD4 T cells that had been selected in some of the readily available mice that expressed a single MHCII protein, IAb, IAf and IAs (Mathis et al., 1983). Transgenic mice that expressed either DOβWT or DOβ48A and no other TCRβ were crossed such that they each expressed one of these MHCII alleles. The numbers of mature CD4 and CD8 T cells in the thymuses of the H2 b, f or s strains of mice were measured. As predicted by our previous data using retrogenic mice (Scott-Browne et al., 2009), the numbers of mature CD4 or CD8 thymocytes in mice expressing DOβ48A were much lower than those in mice expressing DOβWT (**Fig. 1A**). This was true regardless of the MHC allele on which the cells were selected. Thus the TCRβ 48A for 48Y substitution affects MHC interactions regardless of the MHC class or allele, as we have previously predicted (Scott-Browne et al., 2009). The difference in numbers of mature T cells between mice expressing DOβWT and DOβ48A was less marked in peripheral lymph nodes than in the thymus (**Fig.1B),** probably because of increased homeostatic expansion, as exemplified by the increased percentages of CD44hi T cells amongst those few that could mature in DOβ48A mice (**Fig. 1C**).

There were more mature CD4 than CD8 T cells in the lymph nodes of mice expressing DOβWT and H2b (**Fig 1D)**. The effect was much less marked in mice expressing H2f or H2s. The phenomenon may be due to the fact that DOβWT was found in a TCR that reacts with IAd or IAb plus a foreign peptide (OVA 327-339) and not from an MHCI-reactive TCR (White et al., 1983). The bias towards CD4 versus CD8 T cells in H2b animals was not manifest in lymph node cells bearing DOβ48A and was, indeed, reversed in animals expressing that TCRβ and H2f or H2s.

The TCRα confers a bias towards reactivity with the selecting MHC allele

Mature T cells do not usually react detectably with self MHC alleles plus self peptides, the reactivity that presumably allowed their positive selection in the thymus. However, the potential inadequacies of DOβ48A allowed us to test whether or not the TCRαs that, on mature T cells, paired with it did indeed react preferentially with the MHCII allele on which they were positively selected. We guessed that introduction of the more prominently MHC-reactive DOβWT chain into DOβ48A T cells might reveal the underlying reactivity of the TCRα sequences in these T cells for various MHC alleles. Thus we isolated CD4 T cells from mice expressing the DOβ48A transgene, stimulated them with anti-TCR, transduced them with a GFP+ retrovirus expressing the DOβWT chain, and tested the ability of the transductants to react with cells expressing different alleles of MHC. CD69 expression was used as a marker of activation. Non transduced (GFP-) cells in the same cultures were used as controls. In no case did the nontransduced cells show a significant response. However, some of the DOβWT transduced cells responded. Notably, the percentage of the transduced cells that responded to challenge was always greatest if the antigen presenting cells expressed the MHC allele on which the T cells were positively selected (**Fig. 2**). For example, T cells from a DOβ48A H2b mouse, after transduction to express DOβWT, were most likely to react with H2b presenting cells and DOβWT transduced cells from DOβ48A H2s mice reacted only with challenge cells expressing H2s. These experiments show that the TCRα chain that pairs with the transgenic DOβ48A does indeed contribute to the preference of CD4 T cells to react with peptides bound to the MHCII allele involved in positive selection.

Expression of only one TCRβ chain limits the numbers of TCRα sequences that can participate in positive selection

Because allelic exclusion of the TCRα locus is not perfect (Malissen et al., 1992), mature T cells may express two functional TCRα proteins. To be sure that the TCRα chains analyzed in our experimental mice were actually those involved in positive selection of the cells bearing them, we crossed the DOβWT or DOβ48A transgenic, TCRβ-/- mice with TCRα-/- TCRβ-/- animals of each MHC haplotype to generate animals that were DOβWT or DOβ48A transgenic, TCRβ-/-, TCRα+/-. Naïve CD4 T cells were isolated from the lymph nodes of these animals and cDNA coding for their TCRαs were sequenced as previously described (Silberman et al., 2016). PCR and sequencing errors in the germ line encoded portions of these sequences were corrected as described in the Materials and Methods section. To deal with possible sequencing errors in the non germ line encoded portions of CDR3α, sequences that occurred only once in any given sequencing run were eliminated from further analysis. In fact this decision affected the conclusions of all the experiments show below only slightly. Conclusions from analyses that included all sequences, or that eliminated sequences that occurred with the lowest 5% frequency in each sample were similar (data not shown).

Others have previously reported that the T cells in mice expressing a single TCRβ chain have a limited repertoire of TCRα chains by comparison with WT animals (Fukui et al., 1998). To find out whether this applied to CD4 T cells expressing the DOβWT or DOβ48A TCRβ we constructed species accumulation curves for TCRα sequences in B6 and the TCRβ transgenic animals. These were performed by combining the TCRα sequences from all mice of the same genotype (**Fig. 3**) or by plotting the TCRα sequences for individual mice of each genotype (**Supplementary Fig. 1**). Species accumulation curves show that the total number of TCRα sequences we could detect on naïve CD4 T cells in the TCRβ transgenic animals ranged from less than 5,000 to a maximum of about 30,000. These numbers are less than those found in CD4 naïve T cells from B6 mice, which we found to be similar in number to those found on

mouse CD8 T cells (Genolet et al., 2012), >than 10^5 in number (**Fig. 3**). The numbers of TCRα sequences that could partner with DOβWT in selection of CD4 T cells varied considerably with the selecting MHCII allele. More than 4 times more TCRα sequences were apparent in mice expressing IAb versus IAs (**Fig. 3B-D**), perhaps because the TCR from which DOβWT is derived can be selected by IAb (Liu et al., 1996). This effect of MHC allele on the numbers of selected TCRα chains was not evident in animals expressing DOβ48A. Notably, the numbers of different TCRα chains associated with DOβ48A was lower than those associated with DOβWT regardless of the MHCII allele involved, possibly because of the extra demands imposed on TCRα chains by the inadequate TCRβ chain lacking an important MHC contact residue, Y48 (Scott-Browne et al., 2009).

Perhaps the real surprise in these results is how many TCRα sequences can partner with a single TCRβ and participate successfully in positive selection since work in humans and mice suggest that, on peripheral T cells, each TCRβ partners with only about 5-25 different TCRαs (Arstila et al., 1999, Casrouge et al., 2000, Venturi et al., 2011, Li et al., 2016).

Expressed TCRα sequences are strongly influenced by the selecting MHC allele and partner TCRβ.

We compared the frequency with which particular TRAV/TRAJ/CDR3α amino acid sequences, i.e. the entire TCRα sequences, occurred in the various strains of mice. Data of this type can be compared in several ways. The data can be analyzed to find out whether a particular TRAV/TRAJ/CDR3 sequence occurs in each sample, regardless of how often it appears in the set (comparison of unique sequence use). In this case, Jaccard similarity coefficients can be used to measure the similarity between samples. On the other hand, use of particular TRAV/TRAJ/CDR3 sequences can be compared taking into account the number of times a

particular combination occurs. In this case Anne Chao Jaccard abundance based indices

(Chao et al., 2012) are an appropriate statistical tool. Both methods were used in the

comparisons shown in **Fig. 4**. Jaccard analyses showed that the same combination of

TRAV/TRAJ/CDR3 sequences were likely to appear in samples from mice of the same TCRβ

and MHC genotype but were very unlikely to be shared with the T cells from mice expressing a

different MHC allele (**Fig. 4A**). This was just as apparent when the abundance with which the

sequences were expressed was taken into account (**Fig. 4B**). Thus these data show that, given

a single TCRβ, the TCRα sequences that can participate in positive selection are dramatically

affected by the selecting MHCII allele. Moreover, the fact that the values of the Anne Chao

Jaccard analyses for mice of the same MHCII allele are much larger than those of the Jaccard

analyses shows that sequences that appear frequently in one mouse of a given genotype are

more likely to be found in other mice of the same type. Such a result is a manifestation of the

fact that some sequences were repeated many times in all mice of a given MHCII type, whereas

other sequences were rare. This uneven and certainly non-Poissonian distribution of TCR

sequences has been observed before (Correia-Neves et al., 2001, Fazilleau et al., 2005,

Freeman et al., 2009). The phenomenon was not necessarily caused by expansion of single

clones of T cells since, even in sets in which the CDR3α amino acid sequences were identical,

the DNA sequences were not necessarily all the same (data not shown).

Similar analyses were applied to samples from mice in which the selecting MHCII allele was

identical, but the partner TCRβ differed. The data in **Figs. 4C and D** show that the selected

TCRα chains depended on the partner TCRβ, even when the selecting MHC allele was

identical. Close inspection revealed, however, that there was slightly more overlap if the co-

selected TCRβs were different but the selecting MHCII alleles were identical than if the reverse

were true, that is the co-selected TCRβs were the same but the selecting MHCII alleles were different (**Supplementary Fig 2**) (Fink and Bevan, 1978).

A few sequences appear in at least one mouse of each haplotype. For example, 16 sequences appear in DOβWT mice expressing MHCII b, f or s (but not in any DOβ48A animals) (data not shown). Such sequences might belong to yet undiscovered types of T cells that express an invariant TCRα, like iNKT cells or MAIT cells(Chandra and Kronenberg, 2015, Gapin, 2009). We think this is unlikely to be true because, in the complete naïve CD4 T cell sequences, we did not consistently find the sequences of the iNKT cell or MAIT cell TCRαs. Probably this was because the cells bearing the iNKT cell or MAIT cell invariant TCRαs were in the activated/memory T cell populations, which were not examined in our experiments. Were there to be an undiscovered T cell subset bearing another invariant TCRα it would presumably also be in the activated/memory T cell population and therefore not included in our assays.

These data show that positive selection acts on CD4 T cell precursors, via the action of the expressed MHCII allele on particular TCRα/TCRβ pairs.

TRAV usage depends on the selecting MHCII haplotype and the partner TCRβ chain.

In order to find out which element(s) of TCRα determine MHC allele specificity we analyzed each element separately using data from the experiments described above. Others have previously reported that certain TRAVs are used more frequently by CD4 versus CD8 T cells or in mice expressing particular alleles of MHC (Jameson et al., 1990, Pircher et al., 1992, Sim et al., 1996, Simone et al., 1997, Merkenschlager et al., 1994). TRAV rearrangements occur in thymocytes after the cells have rearranged their TCRβ genes (Lindsten et al., 1987). Thus, the frequency with which TRAVs appear on mature naïve CD4 T cells is predicted to depend on a

number of issues, the ease with which the TRAV gene can rearrange (Chen et al., 2015), its ability to pair with the preexisting TCRβ expressed in the cell (Vacchio et al., 1993) and the ability of the TCRα/TCRβ pair to participate in positive, but not negative, selection on the MHCII protein expressed in the thymus.

We first tested whether the expressed MHC haplotype might unexpectedly affect the nature of TRAVs expressed on preselection thymocytes. As shown in **Fig. 5A** and **Supplementary Fig 3**, TRAV usage on preselection thymocytes was similar, regardless of the MHC allele in the donor animal or the coexpressed TCRβ(s). There were, however, some interesting aspects of TRAV use on preselection thymocytes (**Supplementary Fig 3).** TRAVs whose genes are most proximal to the TRAJ locus (TRAVs 17-21) were frequently rearranged, as predicted by previous studies (Villey et al., 1996, Shih et al., 2011, Genolet et al., 2012) (**Supplementary Fig 4**) (note that the TRAVs are arranged by family and not by position in the TRAV locus). However, in preselection thymocytes we also observed frequent rearrangements involving TRAV 1 and members of the TRAV 3, 6, 7, 10, 11 and 14 families. In the cases of the TRAV families the frequently rearranged TRAVs were not always those which are most proximal to the TRAJ locus. For example, amongst the TRAV7 family, the most frequently rearranged member was TRAV7-2D, one of the family members that is furthest from the TRAJ locus, whereas the close relative of TRAV7-2D, TRAV7-2A, was not frequently rearranged. This suggests that chromatin structure, promoter accessibility and use by rearranging processes also play a role in TRAV rearrangements (Chen et al., 2015).

TRAV use by T cells from mice of the same genotype was very similar **(Fig. 5 B-G, Supplementary Fig. 4 and 5**). However, the selecting MHC allele affected the frequency with which different TRAVs were expressed on mature naïve CD4 T cells (**Fig 5** and

Supplementary Figs. 4 and 5). For example members of the TRAV5 family were used to some extent by CD4 T cells selected on IAb, but not by cells selected on IAf or IAs (compare **Figs. 5B, C with Figs. 5D-G**). On the other hand, CD4 T cells in DOβWT H2s mice were alone in their use of TRAV17 **(Fig 5 B, D and F)**. Differential use of TRAVs was much more marked in DOβ48A mice and illustrated the TRAV preferences of mice selected on different MHCII alleles more strikingly. For example, DOβWT cells selected on IAs used most members of the TRAV6 family whereas DOβ48A expressing cells selected on the same MHC allele used, of the TRAV 6 family, almost entirely TRAV6-5D and TRAV6-7DN and also used more frequently than T cells selected on other MHCII alleles, members of the TRAV16 family. Perhaps this reflects a greater need for basic amino acids in TRAV CDR1 and CDR2 for selection of H2s with DOβ48A as a partner, since, of the TRAV6 family, TRAVs6-5D and TRAV6-7DN (and TRAV6-5A and TRAV6-7DN) have a total of two basic amino acids in these elements whereas other members of the family have none. Likewise all expressed members of the TRAV16 family contain 2 or 3 basic amino acids in their CDR1 and CDR2 segments. This narrowing in TRAV choice by DOβ48A cells may reflect the increasing demands for selection imposed by the absence of the tyrosine at position 48 of the TCRβ chain.

Our analyses are based on a method in which cDNAs from individual mouse T cells are amplified simultaneously with a reverse TRAC oligo and oligos built to match each TRAV family (see Methods Section). Therefore the differences in TRAV discovery could be due to more efficient PCR amplification of some TRAV genes than others. However, since the efficiencies of detection will be similar between members of the same family, it is legitimate to compare the frequency of rearrangement between different members of the same family, or the frequency of use of the same TRAV in mature T cells selected on different MHCs or with different TCRβ partners (see below).

TRAJ usage depends on the selecting MHCII haplotype and the partner TCRβ chain.

TRAJ use by preselection thymocytes was similar regardless of the selecting MHC haplotype or co-expressed TCRβ (**Fig. 6A**). TRAJ use by naïve CD4 T cells from B6 mice was fairly uniform across the locus **(Fig. 6B).** Unexpectedly, however, and in contrast to preselection thymocytes and naïve CD4 T cells from B6 mice, TRAJ use by T cells from mice expressing a single TCRβ was much more uneven and tended towards TRAJs whose genes were distal to the TRAV locus (**Fig. 6 B-L**). Regardless of MHC allele, CD4 T cells in DOβWT animals used TRAJ21 most frequently (**Fig. 6 D-F**). The reasons for this bias are unknown. TRAJ21 contains a tyrosine at or near the contact point with MHC but other TRAJs have a tyrosine similarly situated and they are not overexpressed. Moreover TRAJ21 is not overexpressed in T cells expressing DOβ48A, T cells that might be expected to be even more readily selected with an added tyrosine (Scott-Browne et al., 2009). The bias towards use of distal TRAJ genes was even more marked in animals expressing DOβ48A. In these mice TRAJ 9, TRAJ 12 and TRAJs 9,15 and 31 dominated in H2b, H2f and H2s mice respectively (Fig. 6 G-I). Pairwise comparisons between different mice are shown in **Supplementary Fig. 6 and DESeq 2 analyses are in Supplementary Fig 7.**

We do not know why the distal TRAJ genes were preferred in mice in which the TCRα repertoire was limited by the presence of a single TCRβ. In another study with a fixed TCRβ chain, a bias towards proximal TRAJs was noted with TRAV17, a TRAV that is close to the TRAJ locus (Casanova et al., 1991). The same publication described biases, depending on MHC allele towards use of Type 1 (G rich) or type 2 TRAJs. These explanations don't apply here, since in the experiments presented here TRAV expression was not particularly biased towards the distal TRAVs and the used TRAJs don't fall particularly into Types 1 or 2. It is

possible that the choice is related to the DOβ chain itself. Alternatively it maybe that, because it is difficult for thymocytes expressing a single TCRβ to find a TCRα that can pair with the TCRβ and contribute to positive selection, multiple TCRα rearrangements have to occur in each thymocyte before a suitable TCRα partner is found. This will inevitably drive expressed TCRαs towards use of the distal TRAJs, although why these should satisfy the demands of positive selection more frequently than the proximal TRAJs do, is not obvious, at least from their amino acid sequences. It has recently been reported that prolonged expression of RAG protects cells, to some extent, from death (Karo et al., 2014). If the thymocytes in TCRβ transgenic mice have to express RAG for a longer time to find a suitable TCRα partner, then the prolonged expression of RAG needed for the multiple rearrangements required to access the TRAC proximal TRAJs might preferentially allow survival of the thymocytes in which this prolonged expression has occurred. Preliminary analyses of the naïve CD4 T cells in the various mice did not, however, suggest that the T cells in the TCRβ transgenic mice were more resistant to death than the equivalent cells in B6 animals**.**

Thus overall, like the use of TRAVs, use of TRAJs depended on both the MHC haplotype and TCRβ present in the animals.

Expressed CDR3α sequences are strongly influenced by the selecting MHC allele and partner TCRβ.

Because of the removal or introduction of bases when TRAVs rearrange to TRAJs, CDR3α protein sequences can vary in the number of amino acids they encode between the conserved C terminal cysteine of the TRAV and the conserved phenyl alanine/leucine/tryptophan glycine pair in the TRAJ region. We compared the average lengths of CDR3αs and the predicted number of N region bases between mice expressing the same TCRβ and different MHCII

alleles. CD4 T cells selected on IAb had significantly shorter CDR3α lengths and fewer N region bases than their counterparts selected on IAf or IAs (**Supplementary Fig. 8**).

The analyses shown in Fig 4 compared the frequencies with which entire TCRα sequences appeared under different selecting circumstances. We also analyzed how often particular CDR3α sequences are found in mice that differed in the selecting MHC allele or in the co-expressed TCRβ, using, again, Jaccard or Anne Chao Jaccard analyses to compare particular sequences without or with taking into account the abundance with which they occurred. Data comparing the occurrence of CDR3α protein sequences between mice that expressed the same or different MHCII alleles are shown in **Fig. 7A and B**, and data comparing the occurrence of CDR3α protein sequences between mice expressing the same MHCII but different partner TCRβs are shown **in Fig. 7C and D**. The results were similar to those obtained when comparing the entire TCRα sequences. The expressed CDR3α sequences in mice with a particular MHCII allele were very unlikely to be found in CD4 T cells of mice expressing a different MHCII allele, even if the co-selected TCRβ were the same in the mice (**Figs 7A and B**). Likewise, CDR3α sequences co-selected with a particular TCRβ were unlikely to be shared with those co-selected with a different TCRβ, even if the selecting MHCII allele were the same.

The first few amino acids of CDR3α (defined as the stretch between the last C of the TRAVs and the conserved F/W/L G sequence of the TRAJs) are encoded by the TRAVs themselves. Likewise, the last few amino acids of CDR3α are encoded by the TRAJs. Therefore the fact that the CDR3α sequences are controlled by the MHCII allele on which they were selected might have been, to some extent, dictated not by the non germline encoded amino acids in CDR3α but rather by the TRAV encoded amino acids downstream of the cysteine at the C terminal end of the TRAVs or by the TRAJ encoded amino acids upstream of their conserved

TRAJ F/W/L. This problem applies particularly to the use of TRAVs since TRAV CDR1 and

CDR2 amino acids may contact MHCII and thereby contribute to thymic selection whilst also

dictating the first few amino acids of the accompanying CDR3 region.

We therefore checked whether CDR3α sequences associated with particular TRAV/TRAJ pairs

differed between T cells selected on different MHCII alleles or associated with different TCRβs.

Only a few of the possible TRAV/TRAJ pairs were present in sufficient numbers in all of the

mice to be compared, so only a few such comparisons could be made. Examples of such

comparisons are shown in **Fig. 8**. Summaries combining all allowable results (in which all the

TRAV.TRAJ combinations to be compared included least 5 different CDR3 sequences/mouse)

are shown in **Supplementary Fig. 9**. T cells expressing DOβWT and the same TRAV and

TRAJ combinations, but selected on different MHCII alleles or with different TCRβs clearly had

CDR3α sequences that were almost completely unique to the selecting MHCII alleles.

A recent study has reported that thymocytes with aromatic/hydrophobic amino acids at the tips

of their CDR3β segments are biased towards MHC reactivity, regardless of the selecting MHCII

allele(Stadinski et al., 2016). The observations in the paper applied to CD4+ CD8+ (double

positive thymocytes) that had been positively, but not negatively, selected, identified by their

expression of CD69, and to regulatory T cells compared to preselection thymocytes. Such cells

have not been examined in the experiments described here, so we cannot tell directly whether a

similar observation applies to TCRα sequences. On the whole the evidence is that cells with

aromatic amino acids at positions 6 and 7 on CDR3α are not particularly eliminated by clonal

deletion in the thymus (data not shown). Nevertheless, we evaluated individual amino acids

that would probably be at the tips of CDR3αs in CDR3α of different lengths. The results show

MHCII allele and TCRβ specific selection for particular amino acids and also changes in amino

acid preference at CDR3α positions depending on the length of the CDR3 (**Supplementary Figure 10)**. For example, arginine was very frequently used at position 4 in 12 amino acid long CDR3αs selected by IAs with DOβ48A, and similarly over selected at position 5 in 14 amino acid long CDR3α☐s selected on IAf with DOβWT, but much less frequently used by other MHC selection, TCRβ, CDR3α length combinations. Phenyl alanine was only used with evident frequency at position 5 in 14 amino acid-long CDR3αs selected on IAb with DOβWT. Apart from the phenyl alanine result there was no particular enrichment for aromatic amino acids at these tips.

Discussion

It has long been known that T cells bearing αβTCRs are biased towards recognition of antigenic peptides bound to the allele of MHC to which the T cells were exposed in the thymus(Fink and Bevan, 1978, Zinkernagel et al., 1978, Kappler and Marrack, 1978, Sprent, 1978). This phenomenon, known as positive selection, has been ascribed to a requirement for a low affinity/avidity reaction between the developing thymocyte and MHC proteins to which the cell is exposed in the thymus cortex(Sprent et al., 1988). However, the peptides presented to immature T cells in the thymus are also controlled by the allele of MHC involved. Thus, the MHC allele specificity of positive selection might be dictated by TCR contact with the MHC-engaged peptides rather than the MHC protein itself. If this were the case, positive selection might be dominated by the portion of TCRs that most consistently engages the peptides bound to MHC, CDR3 sequences of TCRs, rather than the germ line encoded TRAVs and TRBVs. This idea is supported by the fact that, in some cases, peptides related to the activating antigen can stimulate positive selection of thymocytes bearing particular TCRs (Ashton-Rickardt et al., 1994, Sebzda et al., 1994, Hogquist et al., 1994, Kraj et al., 2001, Smyth et al., 1998). Moreover MHC proteins that were supposed to differ only in amino acids that bind peptide and that don't

contact TCRs nevertheless were found to differ in their ability to select thymocytes bearing certain TCRs (Nikolic-Zugic and Bevan, 1990).

During positive selection could TCRs detect allelic differences between MHCs directly? Although that MHC amino acids that contact TCRs are quite well conserved (Bjorkman et al., 1987) they do vary somewhat between alleles (Reche and Reinherz, 2003). For the MHCII alleles studied here, the amino acids pointing towards the TCR are, at most positions, uniform, but IA^b differs from IA^f and IA^s with a two amino acid insertion on the surface of its beta chain alpha helix, an insertion that causes an allele specific bulge in this helix. IA^s also differs from the other MHCII alleles we studied with alpha chain H68Y and beta chain R70Q changes and position 72 of the MHCIIα alpha helix is a V in IA^b, an I in IA^f and IA^s. Therefore the solvent exposed residues on the alpha helices of the MHCII proteins themselves could contribute to allele specific positive selection. These amino acids are contact points, not only for the CDR1 and CDR2 loops of TCRs, but also, sometimes for the TCR CDR3 regions. For example, in the structure on a TCR bound to the complex of IA^u bound to a myelin basic protein, CDR3α engages polymorphic amino acids at positions 55 and 81 of the IA^u alpha chain and CDR3β engages polymorphic amino acids at positions 65 of the IA^u alpha chain and positions 67 and 70 of the IA^u beta chain (Maynard et al., 2005).

The findings presented previously (Merkenschlager et al., 1994) and here, demonstrate that the allele of MHCII involved in positive selection affects the frequencies with which TRAV and TRAJ elements are selected and, most dramatically, the CDR3α sequences that appear on mature T cells, as previously indicated for CD8 T cells (Ferreira et al., 2006). The N terminal portion of CDR3α is provided by the TRAV, so the effects of MHCII on TRAV choice by a particular TCRα could actually be due to demands placed on the CDR3α rather than on the

TRAV itself. On the other hand, since the CDR3α sequence is affected in part by its co-expressed TRAV, the demands of positive selection could be entirely on the TRAV, not the CDR3α. We think it likely that all three of the TCRα CDRs can play a role in positive selection. However, clearly CDR3α is involved since the sequences in the center of this element vary depend on the selecting MHCII allele, even if the accompanying TRAV and TRAJ are the same **(Fig. 8**).

Overall, the results strongly suggest that positive selection allele specificity involves recognition of both MHC and peptide (reviewed in (Klein et al., 2014, Vrisekoop et al., 2014). In fact, given the geometry with which TCRs engage their MHC/peptide ligands, it is difficult to imagine that this would not be the case.

The data here also show that the sequence of the TCRβ chain affects the TCRα and CDR3α that can participate in positive selection almost as much as the selecting MHC allele does. Not only does the TCRβ affect which TCRαs and which CDR3αs will be successful, it also determines how many different TCRαs can do the job since, regardless of the MHCII allele involved, fewer TCRαs can be selected with DOβ48A than with DOβWT. These results are similar to those observed earlier that showed that fewer CD4 T cells are selected in DOβ48A-expressing versus DOβWT-expressing mice. Together these suggest that DOβ lacking an important MHC contact amino acid, the "Y" at position 48, places more stringent requirements on TCRα for successful thymus selection (Scott-Browne et al., 2009).

Overall, the results show that the entire TCR sequence plays a role in positive selection. How can this be, given that selection is thought to occur during low affinity reactions? Naively one might have predicted that relatively few TCR-to-MHC/peptide interactions would be needed to

reach the needed energy of interaction and these could be provided by just a portion of the TCR, not the entire molecule as suggested here. Some TCR configurations may interfere with contact with MHC/peptide or prevent the proper engagement of CD4 or CD8. Other TCR configurations may react too strongly with their ligand, leading to negative selection. This idea may apply to up to 70% of all TCRs (Ignatowicz et al., 1996, Stritesky et al., 2013). Competition for selecting ligands may also play a role. Also to be bourne in mind is the fact that here we are observing the consequences of many TCR selection events, some TCRs may be selected based on their TRAVs, others via their TCRβs, with the observed results showing biases by both of these. Nevertheless the detrimental effects of an inappropriate CDR3α cannot be overcome by other elements of TCRα.

There are problems with the notion that the bound peptide is a determinant of MHC allele specific positive selection. Most notably, the fact that mature T cells, after selection on a single MHCII allele bound to a single peptide can respond to peptides that are unrelated in sequence to the selecting peptide (Pawlowski et al., 1996, Ignatowicz et al., 1997, Nakano et al., 1997, Ebert et al., 2009, Lo et al., 2009). Moreover, teleologically, the idea that the selecting peptides in the thymus are the only feature that governs T cell specificity doesn't seem evolutionarily favorable. Such might limit the ability of T cells to respond to foreign peptides that are unrelated to those in the thymus. Nevertheless, self peptides might provide an advantage anyway, by supporting the survival of mature T cells and also, perhaps, T cell responses to unrelated peptides when the self and foreign peptides are presented on the same cells (Kirberg et al., 1997, Wulfing et al., 2002). However, MHC-bound peptides on thymus cortical epithelial cells are not necessarily the same as those on peripheral cells (Honey et al., 2002, Murata et al., 2007) so this advantage may not be available for all T cells.

In the studies presented here, the total number of TCRαs that can be selected with a single TCRβ ranges between about 4,600 and 30,000, depending on the selecting MHCII allele and partner TCRβ (Fukui et al., 1998). Are these numbers surprisingly low or high? Based on the estimated numbers of TCRβs and TCRαs that appear in the periphery of an individual, it has previously been estimated that each TCRβ chain can be successfully selected with up to 25 different TCRαs (Arstila et al., 1999, Casrouge et al., 2000). Yet here, and in previous studies, it appears that, for a single TCRβ, the number of possible TCRα partners is at least 3 orders of magnitude larger. How to account for this large disparity? Probably it is caused by competition for selection in the thymus, a phenomenon that has been previously demonstrated (Martins et al., 2014, Visan et al., 2006). In a wild type thymus, each of the immature thymocytes is competing with a huge number of others bearing disparate TCRβ and TCRα sequences. In the thymus of a mouse expressing a single TCRβ, the immature thymocytes bear the same large number of TCRαs, and now all those expressing an even approximately suitable TCRα have the opportunity to be positively selected. This idea may be related to the profound bias towards distal TRAJs reported here, and therefore a predicted increased time available for rearrangements for the thymocytes in single TCRβ transgenic animals.

Materials and Methods

Mice

Mice were purchased from the Jackson Laboratory, Bar Harbor ME and subsequently interbred in the Biological Research Center at National Jewish. Plasmids coding for the DO11.10 TCRβ chain (DOβWT) or its mutant, in which the tyrosine at position 48 was replaced by an alanine (DOβ48A) were created, with the human CD2 promoter to drive expression of the genes (White et al., 1983, Greaves et al., 1989). DNAs coding for the promoters and genes were injected into fertilized C57BL/6J (B6) eggs at the Mouse Genetic Core Facility at National Jewish Health. Mice produced from these eggs were crossed with animals lacking functional TCRβ genes (Mombaerts et al., 1991) and with B10.M (H2f) or B10.S (H2s) animals to create animals expressing the transgenic TCRβ genes, no other TCRβ genes and H2b, f or s. By similar intercrosses animals were produced that expressed no functional TCRα or TCRβ genes and H2b, f or s. These animals were intercrossed to give rise to animals expressing either DOβWT or DOβ48A, no other TCRβ genes, TCRα+/- and H2 b, f or s. Animals were subsequently used for analysis if they expressed the TCRα locus derived from B6 rather than B10 animals.

Animals were handled in strict accordance with good animal practice as defined by the relevant national and/or local animal welfare bodies, and all animal work was approved by the National Jewish Health Animal Care and Use Committee (IACUC). The protocol was approved by National Jewish IACUC (protocol number AS2517).

T cell isolation

Cells were isolated from the thymuses, spleens (B6 analyses) or peripheral lymph nodes (DOβWT or DOβ48A analyses) of 6-14 week old mice. CD4 T cells were isolated by negative selection on MACS columns (Milltenyi Biotech). The cells were stained with antibodies

conjugated to a fluorochrome and specific for: Pacific Blue-CD4, Alexa488-TCRβ, PE-B220, PE-TCRδ, PE-CD8, PE-Cy5-CD25, Alexa647-CD44, PE-Cy7-CD62L, purchased from BD Pharmingen or eBioscience or produced in house. The cells were sorted based on their expression of CD4, TCRβ, low levels of CD44 and high levels of CD62L and absence of staining with PE and PE-Cy5. Cells were sorted into staining buffer (BSS, 2% fetal bovine serum, plus sodium azide) by a MoFloXDP (Beckman Coulter Life Sciences or Synergy SY3200 (iCyt) instruments at the National Jewish Health Flow Cytometry Core Facility.

Retroviral infection of T cells.

Retroviruses expressing DOβWT or DOβ48A and green fluorescent protein were produced as described in (Scott-Browne et al., 2009). CD4 T cells were purified, by negative selection on Automax columns, from the spleens and lymph nodes of DOβ48A transgenic, TCRβ-/- mice expressing various MHC alleles. The cells were activated by 24h culture on plates pre-coated with anti-TCRβ (Ham597) and anti-CD28 (37.51) The supernatants were then removed from the plates and replaced with supernatants containing the DOβWT or DOβ48A retroviruses and 8ug/ml polybrene in culture medium. The cells were spun at 2000G in bags containing $10\%CO_2/90\%$air at 37.C for 2h. At this point the medium was replaced with complete culture medium containing 10% fetal bovine serum and cultured for 1d followed by addition of IL-2. Three days later the cells were harvested and challenged as described below.

Assessment of MHC reactivity of transduced T cells

Red blood cell depleted spleen cells from mice expressing various MHC alleles were cultured overnight with IL-4 plus GM-CSF. The cells were then thoroughly washed and used, at a dose of 10^6 cells/well, to stimulate 10^6/well TCRβ transduced CD4+ T cells, prepared as described above. These wells were cultured for 51/2 hours in a final total volume/well of 200ul of CTM.

The cells were then fixed (Permafix), stained and analyzed for expression of CD4 (PerCP anti-CD4), GFP and CD69 (PE anti-CD69).

TCRα sequencing and analysis

RNA was isolated from purified naïve CD4 T cells, PCR'd to expand TCRa sequences and sequenced as described in (Silberman et al., 2016). Post-sequencing analysis was performed to identify the TRAV and TRAJ genes for each sequence along with its corresponding CDR3. TRAV family and subfamily members were assigned based on the IMGT designations with modifications based on our own analysis of expressed TRAV sequences in B6 mice. IMGT has identified two gene duplication events in the B6 TRAV locus, the "original" genes, most of which are closest to the TRAJ locus are designated by their family number and a number indicating their subfamily membership. Here, for ease of analysis, we have added the letter "A" to their designation, eg TRAV01-1A. TRAV subfamily members in the IMGT designated duplicated "D" and new "N" genes we add the letters "D" or "N", eg TRAV07-6D or TRAV07-6N. In some cases the entire nucleotide sequences of subfamily members are identical and, therefore, indistinguishable by our analyses. In these cases the subfamily members are designated to include all possible source genes, eg TRAV06-3ADN or TRAV06-6AD.

Errors occur during sequencing reactions and accumulate as the numbers of sequences acquired increase (Bolotin et al., 2012, Liu et al., 2014). The sequences were all corrected for errors in the TRAV and TRAJ elements, which do not somatically mutate. However, because the amino acids in and flanking the non germ line encoded CDR3 regions could not be corrected, sequences with errors in these elements are bound to appear at some low frequency and cause a gradual rise in the species accumulation curves. To eliminate these misreads we decided to include in our analyses only those TCRα sequences that occurred more than once in each sample. To correct for sequencing errors within the CDR3, the sequences were modified

by replacing erroneous nucleotides with the appropriate germline-encoded nucleotides whenever a discrepancy was observed. Such correction was possible only when a nucleotide difference could be resolved by aligning to the germline TRAV and/or TRAJ genes. To avoid making inappropriate changes to the potentially non germline encoded portions of CDR3α, such corrections were applied only if the change from the germline sequence occurred more than 3 nucleotides before the predicted end of the TRAV genes or more than three nucleotides after the predicted end of the TRAJ gene. Finally, the amino acid usage within the CDR3α was determined for each sequence to identify any patterns in the CDR3 regions in sequences belonging to T cells from one MHC haplotype versus another. All of the analysis was performed using in-house programs developed in Python 2.7.

In order to represent the differential TRAV and TRAJ gene usage in TCRs sequenced from different mouse samples, we used edgeR from the R/Bioconductor package. A threshold of $p <$ 0.05 was used to identify genes that were most significantly differentially expressed between samples.

Euclidean distances for TRAVs and TRAJs were calculated as $\log_2$ transformed counts per 10^4 sequences.

Acknowledgements

Author Contributions

JSB provided the idea for these experiments and made the constructs to produce the transgenic mice, LG provided the idea for these experiments, helped in analyses and helped write the manuscript, JWK wrote software to analyze the TCRα sequences, helped in analyses and helped write the paper, JW sorted T cells, tested the role of TCRα in MHC restriction and performed the PCR reactions to analyze TCRα sequences, EK analyzed the mice, JC, SL and TD helped analyze the data, PM designed the experiments, analyzed the data and wrote the manuscript.

References

ADAMS, J. J., NARAYANAN, S., BIRNBAUM, M. E., SIDHU, S. S., BLEVINS, S. J., GEE, M. H., SIBENER, L. V., BAKER, B. M., KRANZ, D. M. & GARCIA, K. C. 2016. Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat Immunol,* 17**,** 87-94.

ARCHBOLD, J. K., ELY, L. K., KJER-NIELSEN, L., BURROWS, S. R., ROSSJOHN, J., MCCLUSKEY, J. & MACDONALD, W. A. 2008. T cell allorecognition and MHC restriction--A case of Jekyll and Hyde? *Mol Immunol,* 45**,** 583-98.

ARSTILA, T. P., CASROUGE, A., BARON, V., EVEN, J., KANELLOPOULOS, J. & KOURILSKY, P. 1999. A direct estimate of the human alphabeta T cell receptor diversity. *Science,* 286**,** 958-61.

ASHTON-RICKARDT, P. G., BANDEIRA, A., DELANEY, J. R., VAN KAER, L., PIRCHER, H. P., ZINKERNAGEL, R. M. & TONEGAWA, S. 1994. Evidence for a differential avidity model of T cell selection in the thymus. *Cell,* 76**,** 651-63.

BAKER, B. M., SCOTT, D. R., BLEVINS, S. J. & HAWSE, W. F. 2012. Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism. *Immunol Rev,* 250**,** 10-31.

BARTON, G. M., BEERS, C., DEROOS, P., EASTMAN, S. R., GOMEZ, M. E., FORBUSH, K. A. & RUDENSKY, A. Y. 2002. Positive selection of self-MHC-reactive T cells by individual peptide-MHC class II complexes. *Proc Natl Acad Sci U S A,* 99**,** 6937-42.

BJORKMAN, P. J., SAPER, M. A., SAMRAOUI, B., BENNETT, W. S., STROMINGER, J. L. & WILEY, D. C. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature,* 329**,** 512-8.

BOLOTIN, D. A., MAMEDOV, I. Z., BRITANOVA, O. V., ZVYAGIN, I. V., SHAGIN, D., USTYUGOVA, S. V., TURCHANINOVA, M. A., LUKYANOV, S., LEBEDEV, Y. B. & CHUDAKOV, D. M. 2012. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol,* 42**,** 3073-83.

CABANIOLS, J. P., FAZILLEAU, N., CASROUGE, A., KOURILSKY, P. & KANELLOPOULOS, J. M. 2001. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J Exp Med,* 194**,** 1385-90.

CASANOVA, J. L., ROMERO, P., WIDMANN, C., KOURILSKY, P. & MARYANSKI, J. L. 1991. T cell receptor genes in a series of class I major histocompatibility complex-restricted cytotoxic T lymphocyte clones specific for a Plasmodium berghei nonapeptide: implications for T cell allelic exclusion and antigen-specific repertoire. *J Exp Med,* 174**,** 1371-83.

CASROUGE, A., BEAUDOING, E., DALLE, S., PANNETIER, C., KANELLOPOULOS, J. & KOURILSKY, P. 2000. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J Immunol,* 164**,** 5782-7.

CHANDRA, S. & KRONENBERG, M. 2015. Activation and Function of iNKT and MAIT Cells. *Adv Immunol,* 127**,** 145-201.

CHAO, A., CHIU, C. H. & HSIEH, T. C. 2012. Proposing a resolution to debates on diversity partitioning. *Ecology,* 93**,** 2037-51.

CHEN, L., CARICO, Z., SHIH, H. Y. & KRANGEL, M. S. 2015. A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRdelta and TCRalpha repertoires. *Nat Immunol,* 16**,** 1085-93.

COLF, L. A., BANKOVICH, A. J., HANICK, N. A., BOWERMAN, N. A., JONES, L. L., KRANZ, D. M. & GARCIA, K. C. 2007. How a single T cell receptor recognizes both self and foreign MHC. *Cell,* 129**,** 135-46.

CORREIA-NEVES, M., WALTZINGER, C., MATHIS, D. & BENOIST, C. 2001. The shaping of the T cell repertoire. *Immunity,* 14**,** 21-32.

CRUMPACKER, D. B., ALEXANDER, J., CRESSWELL, P. & ENGELHARD, V. H. 1992. Role of endogenous peptides in murine allogenic cytotoxic T cell responses assessed using transfectants of the antigen-processing mutant 174xCEM.T2. *J Immunol,* 148**,** 3004-11.

DAI, S., HUSEBY, E. S., RUBTSOVA, K., SCOTT-BROWNE, J., CRAWFORD, F., MACDONALD, W. A., MARRACK, P. & KAPPLER, J. W. 2008. Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity,* 28**,** 324-34.

DAVIS, M. M. 1985. Molecular genetics of the T cell-receptor beta chain. *Annu Rev Immunol,* 3**,** 537-60.

DEKOSKY, B. J., IPPOLITO, G. C., DESCHNER, R. P., LAVINDER, J. J., WINE, Y., RAWLINGS, B. M., VARADARAJAN, N., GIESECKE, C., DORNER, T., ANDREWS, S. F., WILSON, P. C., HUNICKE-SMITH, S. P., WILLSON, C. G., ELLINGTON, A. D. & GEORGIOU, G. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol,* 31**,** 166-9.

EBERT, P. J., JIANG, S., XIE, J., LI, Q. J. & DAVIS, M. M. 2009. An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. *Nat Immunol,* 10**,** 1162-9.

FAZILLEAU, N., CABANIOLS, J. P., LEMAITRE, F., MOTTA, I., KOURILSKY, P. & KANELLOPOULOS, J. M. 2005. Valpha and Vbeta public repertoires are highly conserved in terminal deoxynucleotidyl transferase-deficient mice. *J Immunol,* 174**,** 345-55.

FERREIRA, C., FURMANSKI, A., MILLRAIN, M., BARTOK, I., GUILLAUME, P., LEES, R., SIMPSON, E., MACDONALD, H. R. & DYSON, J. 2006. TCR-alpha CDR3 loop audition regulates positive selection. *J Immunol,* 177**,** 2477-85.

FINK, P. J. & BEVAN, M. J. 1978. H-2 antigens of the thymus determine lymphocyte specificity. *J Exp Med,* 148**,** 766-75.

FREEMAN, J. D., WARREN, R. L., WEBB, J. R., NELSON, B. H. & HOLT, R. A. 2009. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res,* 19**,** 1817-24.

FUKUI, Y., HASHIMOTO, O., INAYOSHI, A., GYOTOKU, T., SANO, T., KOGA, T., GUSHIMA, T. & SASAZUKI, T. 1998. Highly restricted T cell repertoire shaped by a single major histocompatibility complex-peptide ligand in the presence of a single rearranged T cell receptor beta chain. *J Exp Med,* 188**,** 897-907.

GAPIN, L. 2009. Where do MAIT cells fit in the family of unconventional T cells? *PLoS Biol,* 7**,** e70.

GARCIA, K. C., DEGANO, M., STANFIELD, R. L., BRUNMARK, A., JACKSON, M. R., PETERSON, P. A., TEYTON, L. & WILSON, I. A. 1996. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science,* 274**,** 209-19.

GARCIA, K. C., TEYTON, L. & WILSON, I. A. 1999. Structural basis of T cell recognition. *Annu Rev Immunol,* 17**,** 369-97.

GELLERT, M. 2002. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem,* 71**,** 101-32.

GENOLET, R., STEVENSON, B. J., FARINELLI, L., OSTERAS, M. & LUESCHER, I. F. 2012. Highly diverse TCRalpha chain repertoire of pre-immune CD8(+) T cells reveals new insights in gene recombination. *EMBO J,* 31**,** 4247-8.

GRANDEA, A. G., 3RD & BEVAN, M. J. 1993. A conservative mutation in a class I MHC molecule outside the peptide binding groove stimulates responses to self peptides. *J Immunol,* 151**,** 3981-7.

GREAVES, D. R., WILSON, F. D., LANG, G. & KIOUSSIS, D. 1989. Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell,* 56**,** 979-86.

HOGQUIST, K. A., JAMESON, S. C., HEATH, W. R., HOWARD, J. L., BEVAN, M. J. & CARBONE, F. R. 1994. T cell receptor antagonist peptides induce positive selection. *Cell,* 76**,** 17-27.

HONEY, K., NAKAGAWA, T., PETERS, C. & RUDENSKY, A. 2002. Cathepsin L regulates CD4+ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. *J Exp Med,* 195**,** 1349-58.

HSIEH, C. S., ZHENG, Y., LIANG, Y., FONTENOT, J. D. & RUDENSKY, A. Y. 2006. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat Immunol,* 7**,** 401-10.

HUNIG, T. & BEVAN, M. J. 1981. Specificity of T-cell cones illustrates altered self hypothesis. *Nature,* 294**,** 460-2.

HUNT, H. D., PULLEN, J. K., DICK, R. F., BLUESTONE, J. A. & PEASE, L. R. 1990. Structural basis of Kbm8 alloreactivity. Amino acid substitutions on the beta-pleated floor of the antigen recognition site. *J Immunol,* 145**,** 1456-62.

IGNATOWICZ, L., KAPPLER, J. & MARRACK, P. 1996. The repertoire of T cells shaped by a single MHC/peptide ligand. *Cell,* 84**,** 521-9.

IGNATOWICZ, L., REES, W., PACHOLCZYK, R., IGNATOWICZ, H., KUSHNIR, E., KAPPLER, J. & MARRACK, P. 1997. T cells can be activated by peptides that are unrelated in sequence to their selecting peptide. *Immunity,* 7**,** 179-86.

JAMESON, S. C., KAYE, J. & GASCOIGNE, N. R. 1990. A T cell receptor V alpha region selectively expressed in CD4+ cells. *J Immunol,* 145**,** 1324-31.

KAPPLER, J. W. & MARRACK, P. 1978. The role of H-2 linked genes in helper T-cell function. IV. Importance of T-cell genotype and host environment in I-region and Ir gene expression. *J Exp Med,* 148**,** 1510-22.

KAPPLER, J. W., ROEHM, N. & MARRACK, P. 1987. T cell tolerance by clonal elimination in the thymus. *Cell,* 49**,** 273-80.

KARO, J. M., SCHATZ, D. G. & SUN, J. C. 2014. The RAG recombinase dictates functional heterogeneity and cellular fitness in natural killer cells. *Cell,* 159**,** 94-107.

KIRBERG, J., BERNS, A. & VON BOEHMER, H. 1997. Peripheral T cell survival requires continual ligation of the T cell receptor to major histocompatibility complex-encoded molecules. *J Exp Med,* 186**,** 1269-75.

KLEIN, L., KYEWSKI, B., ALLEN, P. M. & HOGQUIST, K. A. 2014. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol,* 14**,** 377-91.

KRAJ, P., PACHOLCZYK, R., IGNATOWICZ, H., KISIELOW, P., JENSEN, P. & IGNATOWICZ, L. 2001. Positive selection of CD4(+) T cells is induced in vivo by agonist and inhibited by antagonist peptides. *J Exp Med,* 194**,** 407-16.

LI, H. M., HIROI, T., ZHANG, Y., SHI, A., CHEN, G., DE, S., METTER, E. J., WOOD, W. H., 3RD, SHAROV, A., MILNER, J. D., BECKER, K. G., ZHAN, M. & WENG, N. P. 2016. TCRbeta repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. *J Leukoc Biol,* 99**,** 505-13.

LINDSTEN, T., FOWLKES, B. J., SAMELSON, L. E., DAVIS, M. M. & CHIEN, Y. H. 1987. Transient rearrangements of the T cell antigen receptor alpha locus in early thymocytes. *J Exp Med,* 166**,** 761-75.

LIU, C. P., KAPPLER, J. W. & MARRACK, P. 1996. Thymocytes can become mature T cells without passing through the CD4+ CD8+, double-positive stage. *J Exp Med,* 184**,** 1619-30.

LIU, P., LIU, D., YANG, X., GAO, J., CHEN, Y., XIAO, X., LIU, F., ZOU, J., WU, J., MA, J., ZHAO, F., ZHOU, X., GAO, G. F. & ZHU, B. 2014. Characterization of human alphabetaTCR repertoire and discovery of D-D fusion in TCRbeta chains. *Protein Cell,* 5**,** 603-15.

LO, W. L., FELIX, N. J., WALTERS, J. J., ROHRS, H., GROSS, M. L. & ALLEN, P. M. 2009. An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells. *Nat Immunol,* 10**,** 1155-61.

LU, H., SHIMAZAKI, N., RAVAL, P., GU, J., WATANABE, G., SCHWARZ, K., SWANSON, P. C. & LIEBER, M. R. 2008. A biochemically defined system for coding joint formation in V(D)J recombination. *Mol Cell,* 31**,** 485-97.

MALISSEN, M., TRUCY, J., JOUVIN-MARCHE, E., CAZENAVE, P. A., SCOLLAY, R. & MALISSEN, B. 1992. Regulation of TCR alpha and beta gene allelic exclusion during T-cell development. *Immunol Today,* 13**,** 315-22.

MARTINS, V. C., BUSCH, K., JURAEVA, D., BLUM, C., LUDWIG, C., RASCHE, V., LASITSCHKA, F., MASTITSKY, S. E., BRORS, B., HIELSCHER, T., FEHLING, H. J. & RODEWALD, H. R. 2014. Cell competition is a tumour suppressor mechanism in the thymus. *Nature,* 509**,** 465-70.

MATHIS, D. J., BENOIST, C., WILLIAMS, V. E., 2ND, KANTER, M. & MCDEVITT, H. O. 1983. Several mechanisms can account for defective E alpha gene expression in different mouse haplotypes. *Proc Natl Acad Sci U S A,* 80**,** 273-7.

MAYNARD, J., PETERSSON, K., WILSON, D. H., ADAMS, E. J., BLONDELLE, S. E., BOULANGER, M. J., WILSON, D. B. & GARCIA, K. C. 2005. Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity,* 22**,** 81-92.

MAZZA, C., AUPHAN-ANEZIN, N., GREGOIRE, C., GUIMEZANES, A., KELLENBERGER, C., ROUSSEL, A., KEARNEY, A., VAN DER MERWE, P. A., SCHMITT-VERHULST, A. M. & MALISSEN, B. 2007. How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? *EMBO J,* 26**,** 1972-83.

MERKENSCHLAGER, M., BENOIST, C. & MATHIS, D. 1994. MHC control of the naive TCR alpha-chain repertoire. *J Immunol,* 153**,** 3005-13.

MOMBAERTS, P., CLARKE, A. R., HOOPER, M. L. & TONEGAWA, S. 1991. Creation of a large genomic deletion at the T-cell antigen receptor beta-subunit locus in mouse embryonic stem cells by gene targeting. *Proc Natl Acad Sci U S A,* 88**,** 3084-7.

MOSHOUS, D., CALLEBAUT, I., DE CHASSEVAL, R., CORNEO, B., CAVAZZANA-CALVO, M., LE DEIST, F., TEZCAN, I., SANAL, O., BERTRAND, Y., PHILIPPE, N., FISCHER, A. & DE VILLARTAY, J. P. 2001. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell,* 105**,** 177-86.

MUNSON, D. J., EGELSTON, C. A., CHIOTTI, K. E., PARRA, Z. E., BRUNO, T. C., MOORE, B. L., NAKANO, T. A., SIMONS, D. L., JIMENEZ, G., YIM, J. H., ROZANOV, D. V., FALTA, M. T., FONTENOT, A. P., REYNOLDS, P. R., LEACH, S. M., BORGES, V. F., KAPPLER, J. W., SPELLMAN, P. T., LEE, P. P. & SLANSKY, J. E. 2016. Identification of shared TCR sequences from T cells in human breast cancer using emulsion RT-PCR. *Proc Natl Acad Sci U S A,* 113**,** 8272-7.

MURATA, S., SASAKI, K., KISHIMOTO, T., NIWA, S., HAYASHI, H., TAKAHAMA, Y. & TANAKA, K. 2007. Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science,* 316**,** 1349-53.

NAKANO, N., ROOKE, R., BENOIST, C. & MATHIS, D. 1997. Positive selection of T cells induced by viral delivery of neopeptides to the thymus. *Science,* 275**,** 678-83.

NIKOLIC-ZUGIC, J. & BEVAN, M. J. 1990. Role of self-peptides in positively selecting the T-cell repertoire. *Nature,* 344**,** 65-7.

PAWLOWSKI, T. J., SINGLETON, M. D., LOH, D. Y., BERG, R. & STAERZ, U. D. 1996. Permissive recognition during positive selection. *Eur J Immunol,* 26**,** 851-7.

PIRCHER, H., REBAI, N., GROETTRUP, M., GREGOIRE, C., SPEISER, D. E., HAPP, M. P., PALMER, E., ZINKERNAGEL, R. M., HENGARTNER, H. & MALISSEN, B. 1992. Preferential positive selection of V alpha 2+ CD8+ T cells in mouse strains expressing both H-2k and T cell receptor V alpha a haplotypes: determination with a V alpha 2-specific monoclonal antibody. *Eur J Immunol,* 22**,** 399-404.

RECHE, P. A. & REINHERZ, E. L. 2003. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol,* 331**,** 623-41.

REINHERZ, E. L., TAN, K., TANG, L., KERN, P., LIU, J., XIONG, Y., HUSSEY, R. E., SMOLYAR, A., HARE, B., ZHANG, R., JOACHIMIAK, A., CHANG, H. C., WAGNER, G. & WANG, J. 1999. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science,* 286**,** 1913-21.

ROBINS, H. S., SRIVASTAVA, S. K., CAMPREGHER, P. V., TURTLE, C. J., ANDRIESEN, J., RIDDELL, S. R., CARLSON, C. S. & WARREN, E. H. 2010. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med,* 2**,** 47ra64.

RUDOLPH, M. G., STANFIELD, R. L. & WILSON, I. A. 2006. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol,* 24**,** 419-66.

SCOTT-BROWNE, J. P., WHITE, J., KAPPLER, J. W., GAPIN, L. & MARRACK, P. 2009. Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature,* 458**,** 1043-6.

SEBZDA, E., WALLACE, V. A., MAYER, J., YEUNG, R. S., MAK, T. W. & OHASHI, P. S. 1994. Positive and negative thymocyte selection induced by different concentrations of a single peptide. *Science,* 263**,** 1615-8.

SHIH, H. Y., HAO, B. & KRANGEL, M. S. 2011. Orchestrating T-cell receptor alpha gene assembly through changes in chromatin structure and organization. *Immunol Res,* 49**,** 192-201.

SILBERMAN, D., KROVI, S. H., TUTTLE, K. D., CROOKS, J., REISDORPH, R., WHITE, J., GROSS, J., MATSUDA, J. L., GAPIN, L., MARRACK, P. & KAPPLER, J. W. 2016. Class II major histocompatibility complex mutant mice to study the germ-line bias of T-cell antigen receptors. *Proc Natl Acad Sci U S A,* 113**,** E5608-17.

SIM, B. C., ZERVA, L., GREENE, M. I. & GASCOIGNE, N. R. 1996. Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science,* 273**,** 963-6.

SIMONE, E. A., YU, L., WEGMANN, D. R. & EISENBARTH, G. S. 1997. T cell receptor gene polymorphisms associated with anti-insulin, autoimmune T cells in diabetes-prone NOD mice. *J Autoimmun,* 10**,** 317-21.

SMYTH, L. A., WILLIAMS, O., HUBY, R. D., NORTON, T., ACUTO, O., LEY, S. C. & KIOUSSIS, D. 1998. Altered peptide ligands induce quantitatively but not qualitatively different intracellular signals in primary thymocytes. *Proc Natl Acad Sci U S A,* 95**,** 8193-8.

SPRENT, J. 1978. Restricted helper function of F1 hybrid T cells positively selected to heterologous erythrocytes in irradiated parental strain mice. I. Failure to collaborate with B cells of the opposite parental strain not associated with active suppression. *J Exp Med,* 147**,** 1142-58.

SPRENT, J., GAO, E. K., KANAGAWA, O. & WEBB, S. R. 1988. T-cell selection in the thymus. *Princess Takamatsu Symp,* 19**,** 127-36.

STADINSKI, B. D., SHEKHAR, K., GOMEZ-TOURINO, I., JUNG, J., SASAKI, K., SEWELL, A. K., PEAKMAN, M., CHAKRABORTY, A. K. & HUSEBY, E. S. 2016. Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat Immunol,* 17**,** 946-55.

STRITESKY, G. L., XING, Y., ERICKSON, J. R., KALEKAR, L. A., WANG, X., MUELLER, D. L., JAMESON, S. C. & HOGQUIST, K. A. 2013. Murine thymic selection quantified using a unique method to capture deleted T cells. *Proc Natl Acad Sci U S A,* 110**,** 4679-84.

TAN, Y. C., BLUM, L. K., KONGPACHITH, S., JU, C. H., CAI, X., LINDSTROM, T. M., SOKOLOVE, J. & ROBINSON, W. H. 2014. High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin Immunol,* 151**,** 55-65.

TOURNE, S., MIYAZAKI, T., OXENIUS, A., KLEIN, L., FEHR, T., KYEWSKI, B., BENOIST, C. & MATHIS, D. 1997. Selection of a broad repertoire of CD4+ T cells in H-2Ma0/0 mice. *Immunity,* 7**,** 187-95.

VACCHIO, M. S., GRANGER, L., KANAGAWA, O., MALISSEN, B., TOMONARI, K., SHARROW, S. O. & HODES, R. J. 1993. T cell receptor V alpha-V beta combinatorial selection in the expressed T cell repertoire. *J Immunol,* 151**,** 1322-7.

VENTURI, V., QUIGLEY, M. F., GREENAWAY, H. Y., NG, P. C., ENDE, Z. S., MCINTOSH, T., ASHER, T. E., ALMEIDA, J. R., LEVY, S., PRICE, D. A., DAVENPORT, M. P. & DOUEK, D. C. 2011. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol,* 186**,** 4285-94.

VILLEY, I., CAILLOL, D., SELZ, F., FERRIER, P. & DE VILLARTAY, J. P. 1996. Defect in rearrangement of the most 5' TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. *Immunity,* 5**,** 331-42.

VISAN, I., YUAN, J. S., TAN, J. B., CRETEGNY, K. & GUIDOS, C. J. 2006. Regulation of intrathymic T-cell development by Lunatic Fringe- Notch1 interactions. *Immunol Rev,* 209**,** 76-94.

VON BOEHMER, H., TEH, H. S. & KISIELOW, P. 1989. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol Today,* 10**,** 57-61.

VRISEKOOP, N., MONTEIRO, J. P., MANDL, J. N. & GERMAIN, R. N. 2014. Revisiting thymic positive selection and the mature T cell repertoire for antigen. *Immunity,* 41**,** 181-90.

WARREN, R. L., FREEMAN, J. D., ZENG, T., CHOE, G., MUNRO, S., MOORE, R., WEBB, J. R. & HOLT, R. A. 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res,* 21**,** 790-7.

WHITE, J., HASKINS, K. M., MARRACK, P. & KAPPLER, J. 1983. Use of I region-restricted, antigen-specific T cell hybridomas to produce idiotypically specific anti-receptor antibodies. *J Immunol,* 130**,** 1033-7.

WONG, P. & RUDENSKY, A. Y. 1996. Phenotype and function of CD4+ T cells in mice lacking invariant chain. *J Immunol,* 156**,** 2133-42.

WULFING, C., SUMEN, C., SJAASTAD, M. D., WU, L. C., DUSTIN, M. L. & DAVIS, M. M. 2002. Costimulation and endogenous MHC ligands contribute to T cell recognition. *Nat Immunol,* 3**,** 42-7.

ZINKERNAGEL, R. M., ALTHAGE, A., COOPER, S., CALLAHAN, G. & KLEIN, J. 1978. In irradiation chimeras, K or D regions of the chimeric host, not of the donor lymphocytes, determine immune responsiveness of antiviral cytotoxic T cells. *J Exp Med,* 148**,** 805-10.
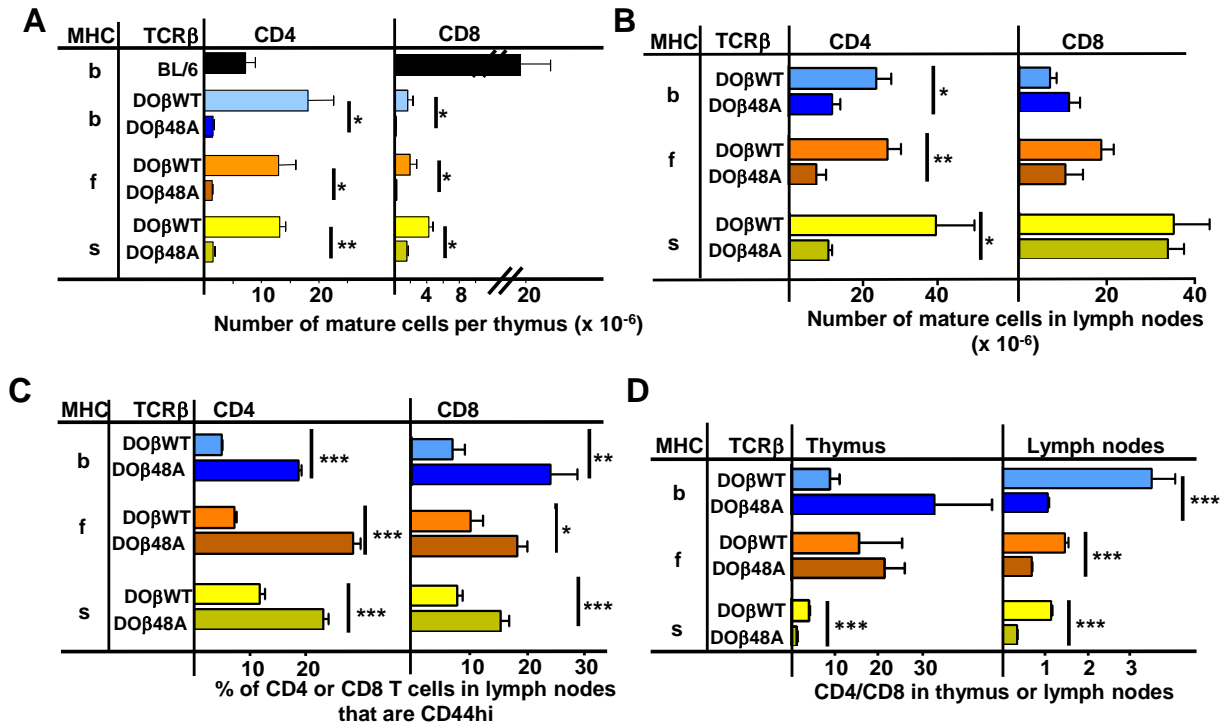
Figure 1. CD4 selection in mice expressing single TCRβ chains and different MHC alleles. Cells were isolated from the thymuses and lymph nodes of mice expressing a single TCRβ, DOβWT or DOβ48A , and different MHC haplotypes and stained for expression of CD4 and CD8 and CD44. Results are the means and standard errors of the mean (SEMs) of three independently analyzed mice expressing the indicated TCRβs and MHC alleles. Student t analyses were used to compare results between the DOβWT and DOβ48A paired samples. *p<0.05, **p<0.01, ***p<0.001.
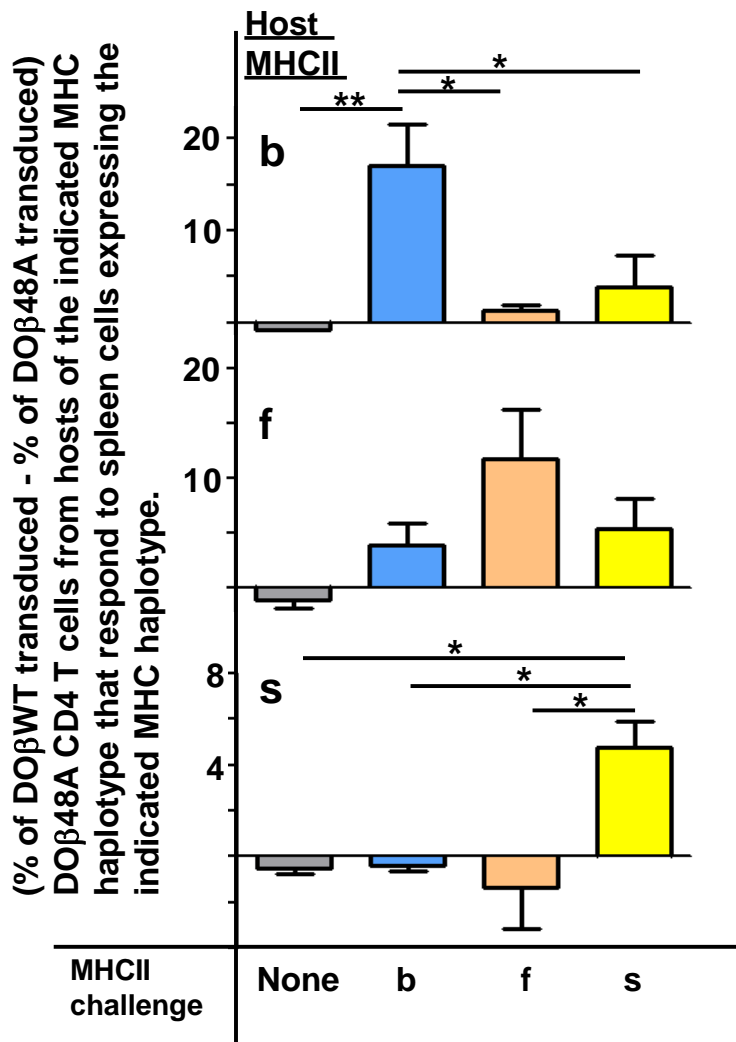
Figure 2. TCRα contributes to the MHCII allele bias of selected naïve CD4 T cells. Naïve CD4 T cells were isolated from the lymph nodes of DOβ48A H2 b, f or s mice and incubated for 2 days in wells coated with anti-TCRβ and anti-CD28. Thus activated, the T cells were spinfected with GFP-expressing retroviruses expressing also DOβWT or DOβ48A. The cells were cultured for a further 2 days and then challenged with spleen cells from mice expressing the indicated MHC alleles, or in the absence of added spleen cells. One day later the cells were stained for expression of CD69. Results were calculated as the (% of GFP+ T cells transduced with DOβWT-expressing retroviruses that were CD69+) – (the % of GFP+ T cells transduced with DOβ48A-expressing retroviruses that were CD69+) in wells containing the same challenge spleen cells. Shown are the average results +/- standard error of the mean (SEM) from 3 independent experiments. *p<0.05, **p<0.01 by one way ANOVA followed by Newman Keuls analyses.
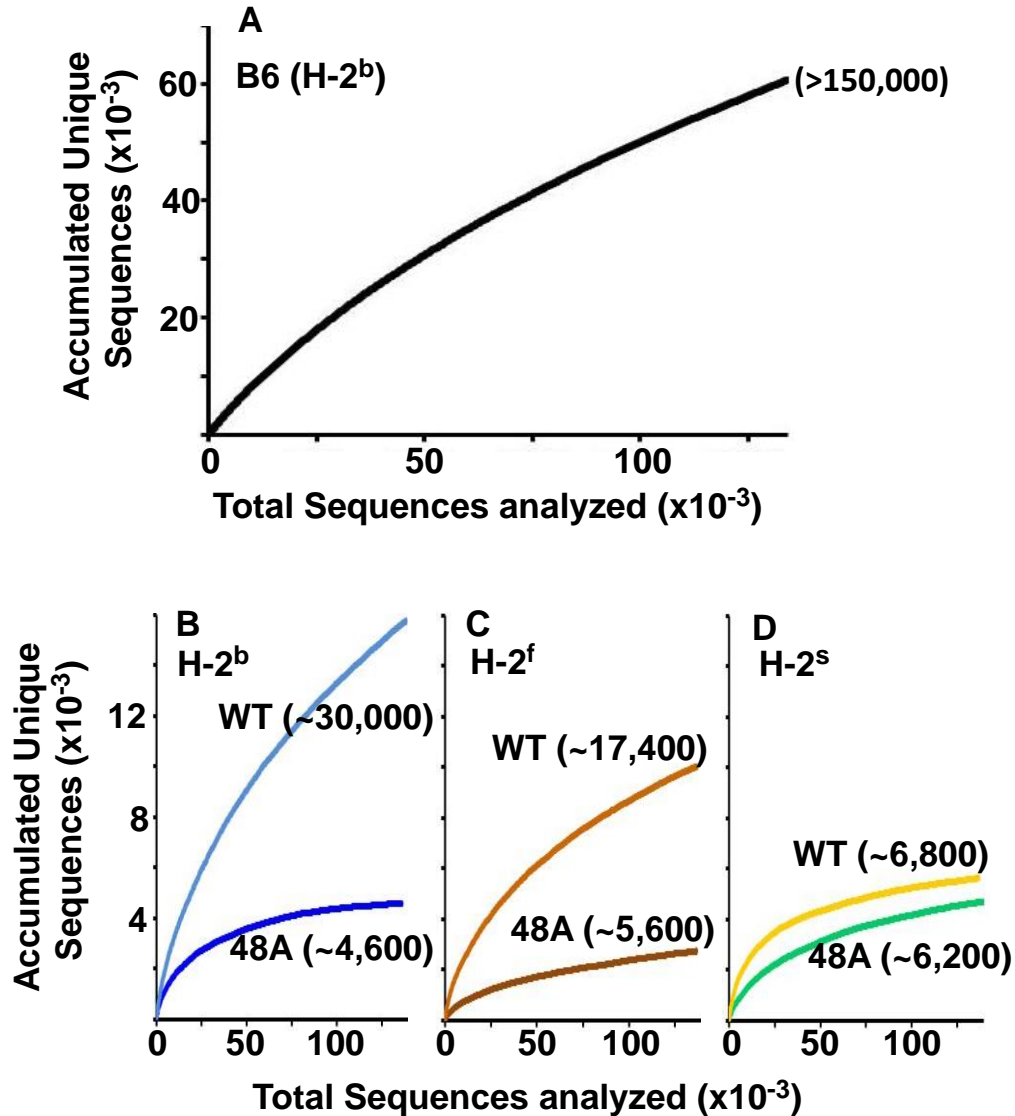
Figure 3. Expression of a single TCRβ chain, DOβWT and, even more markedly, DOβ48A, reduces the number of different TCRα chains that can be positively selected, regardless of the selecting MHCII allele. Naïve CD4 T cells were isolated from the spleens (B6) or lymph nodes of mice expressing MHC b, f or s, single TCRβ chains and heterozygous for expression of functional TCRα chains. Their expressed TCRα chains were sequenced and analyzed with species accumulation curves. Results were combined from 3 independently sequenced data sets from mice of each genotype except for those for H2s DOβWT animals, which were combined from only two independently sequenced animals. Data are shown together with an estimate (bracketed) of the total numbers of different TCRα protein sequences present in the naïve CD4 T cells of each type of mouse.
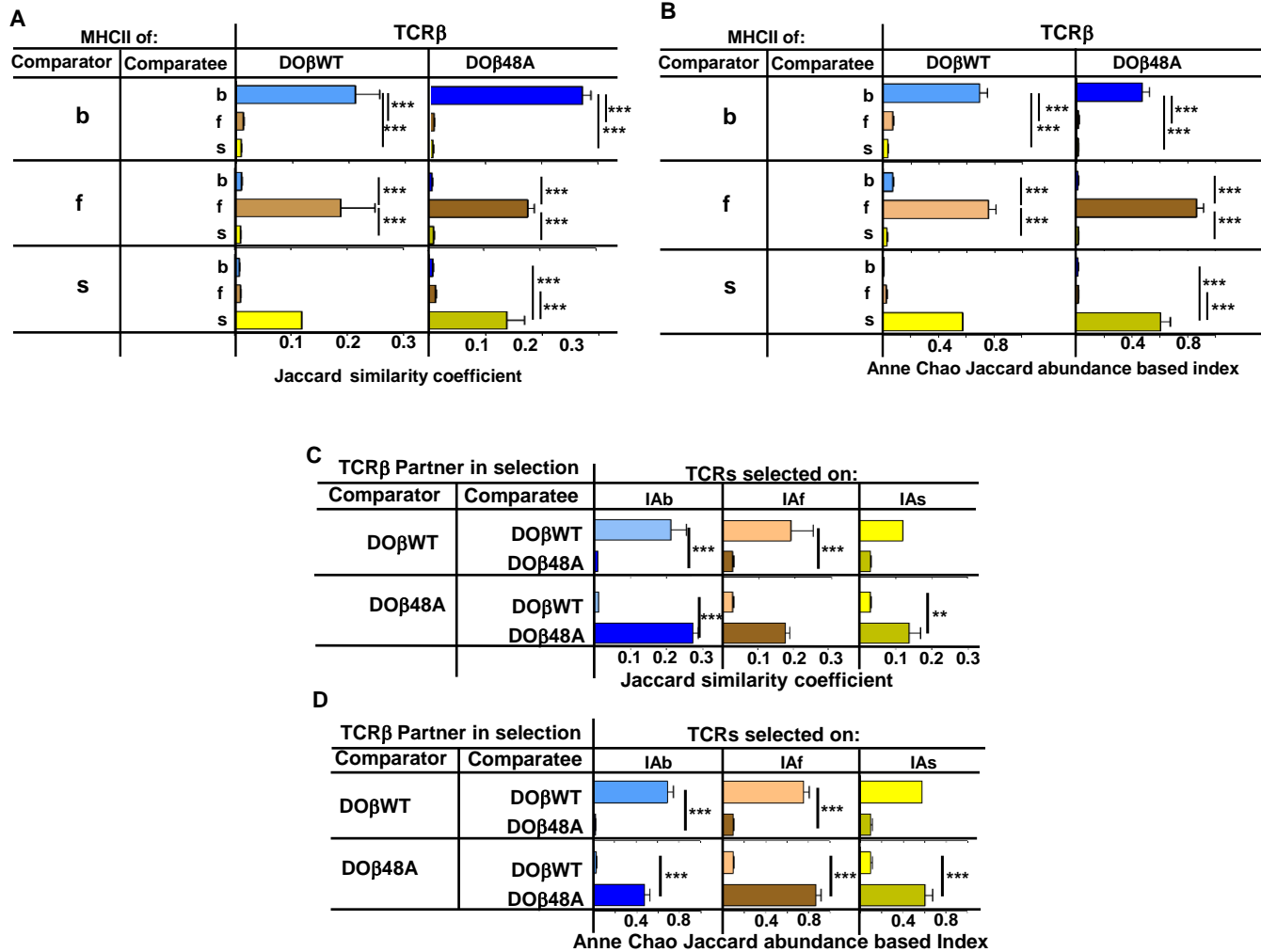
Fig. 4. TCRα sequences on naïve CD4 T cells are determined by the selecting MHCII allele and the co-selected TCRβ. TCRαs on naïve CD4 T cells from the lymph nodes of TCRα+/- mice expressing a single TCRβ and various MHC alleles were sequenced and analyzed as described in Fig. 3. Results are the means and SEMs of three independently sequenced animals of each genotype except for H2s DOβWT animals, of which only two mice were analyzed. ***p<0.001 by one way ANOVA with Newman-Keuls post analysis.
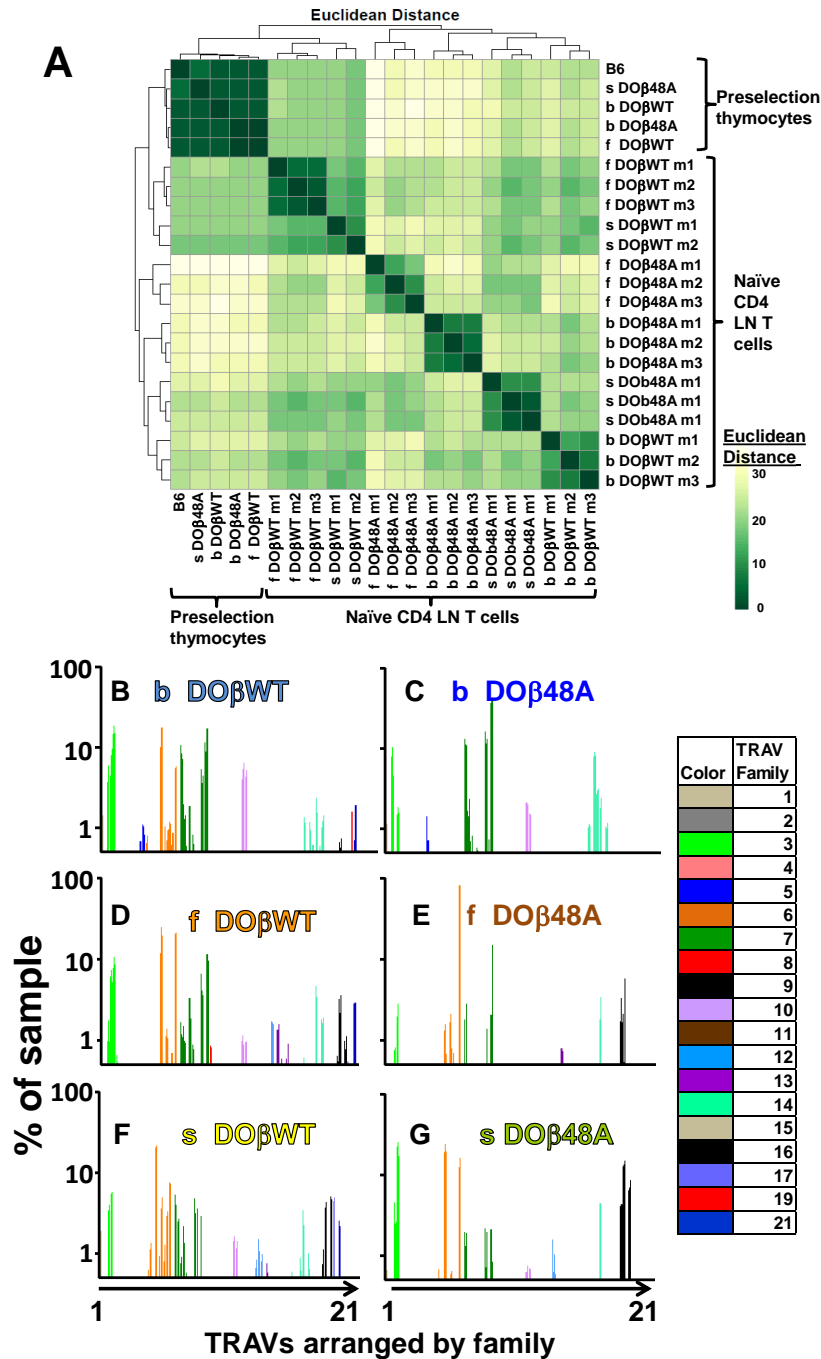
Fig. 5 The frequency with which TRAVs are used on naïve CD4 T cells in TCRβ transgenic mice depends on their selecting MHCII and their partner TCRβ. TCRαs on preselection thymocytes or naïve CD4 T cells from the lymph nodes of TCRα+/- mice expressing a single TCRβ and various MHC alleles were sequenced and analyzed as described in Fig. 3. A. Shown are the Euclidean distances for TRAV use between the data for individual mice. Samples are hierachically ordered. Individual mice of the same genotype are numbered m1-3. B. The average % use of each TRAV in mice expressing the indicated MHCII allele and TCRβ. Results are the means +/- SEMs of 3 identical mice, except for H2s DOβWT animals, for which results are the averages of 2 mice.   TRAVs are ordered by family, not by position on the chromosome.
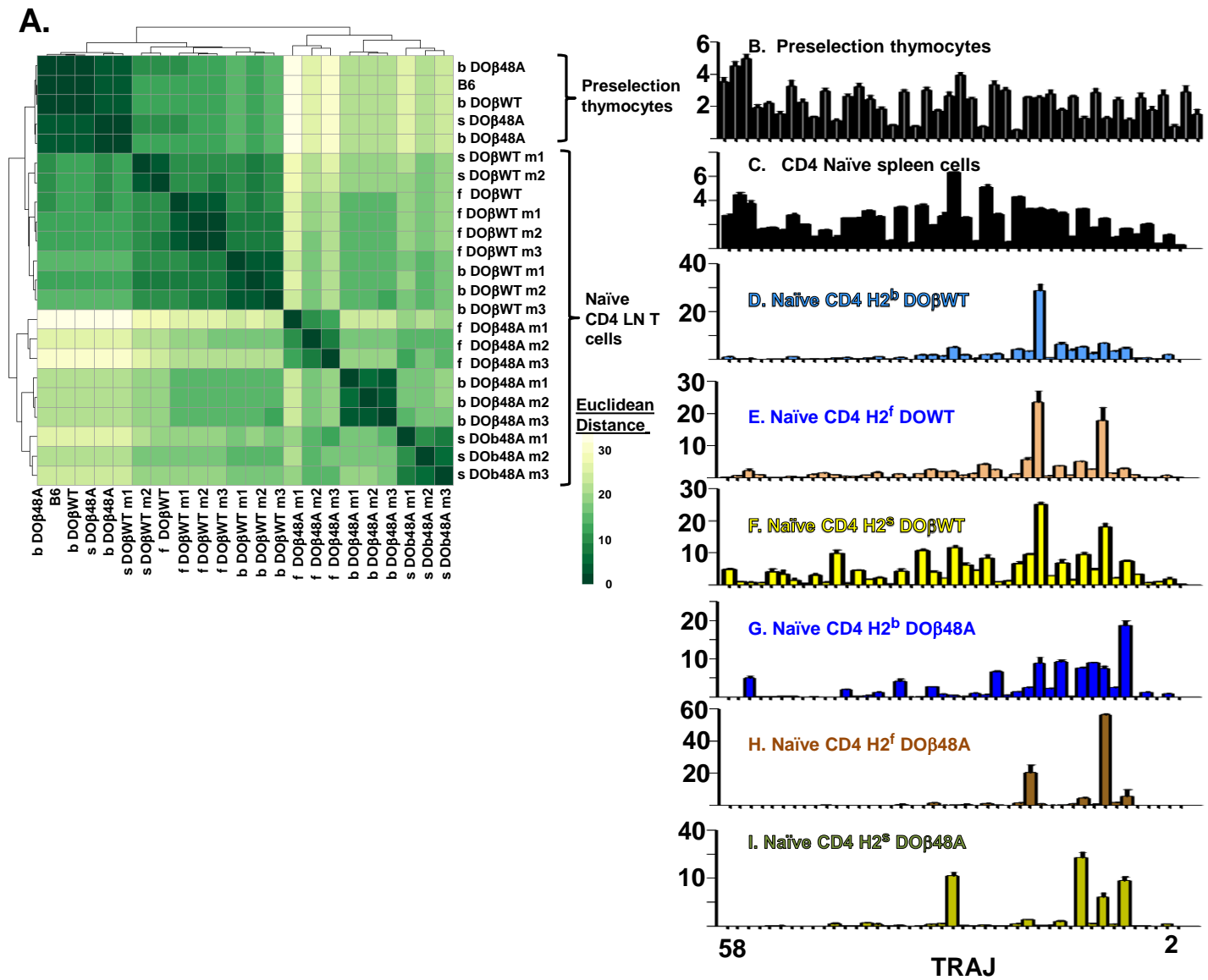
Fig. 6 Naïve CD4 T cells in DOβWT or DOβ48A mice preferentially use TRAJs from the distal end of the TRAJ locus. TCRαs on preselection thymocytes or naïve CD4 T cells from the lymph nodes of TCRα+/- mice expressing a single TCRβ and various MHC alleles were sequenced and analyzed as described in Fig. 3. A. Shown are the Euclidean distances for TRAJ use between the data for individual mice. Samples are hierachically ordered. Individual mice of the same genotype are numbered m1-3. B-L. The % use of each TRAJ in mice expressing the indicated MHCII allele and TCRβ. Results are the means and SEMs of 3 identical mice, except for H2s DOβWT animals, for which results are the averages of 2 identical mice. TRAJs are ordered by position on the chromosome.   Also shown are the means and SEMs of TRAJ use by 5 independently sequenced preselection thymocytes and 3 independently sequenced naïve CD4 T spleen T cells from B6 mice.
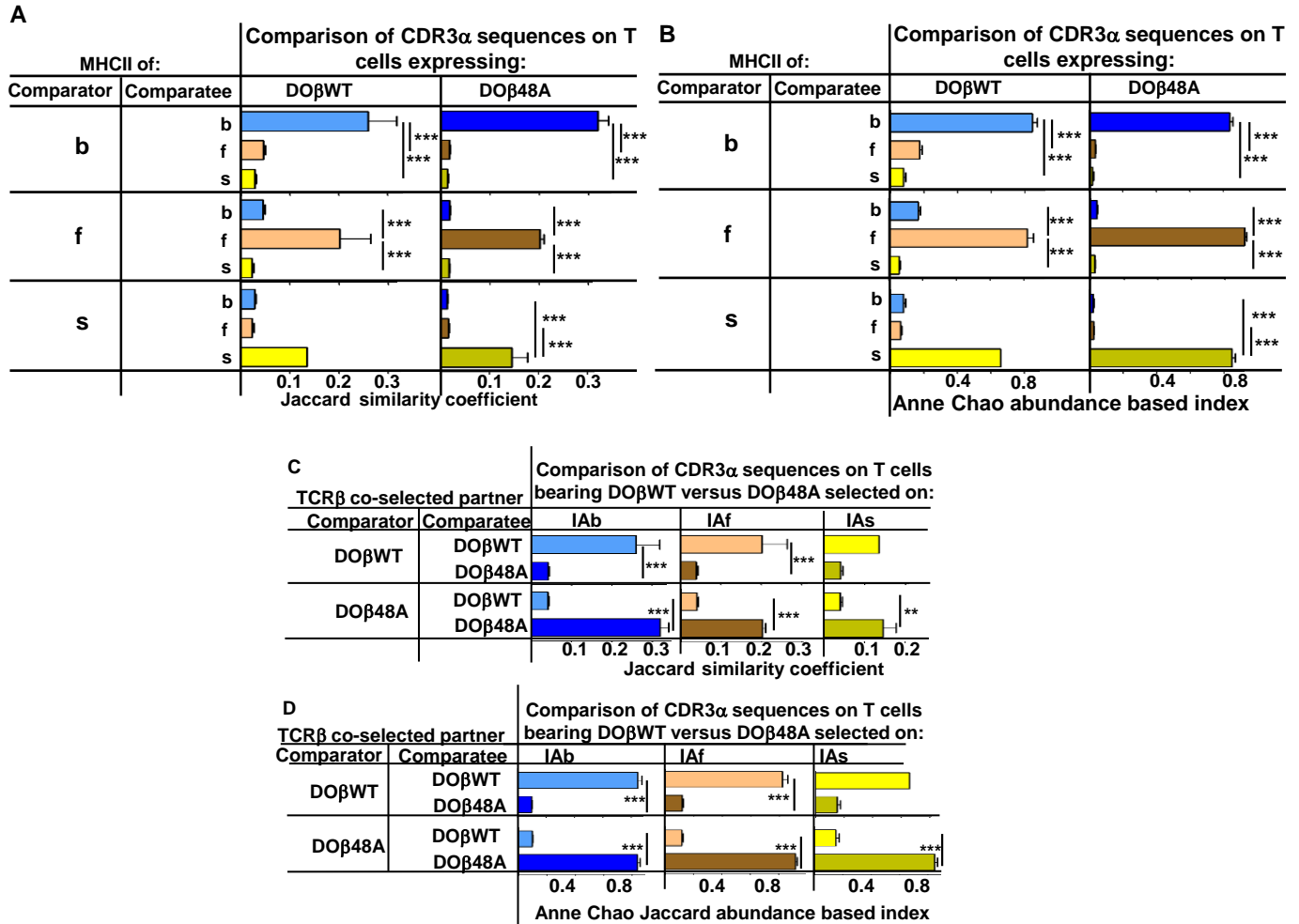
Fig. 7. CDR3α sequences on naïve CD4 T cells are determined by the selecting MHCII allele and the co-selected TCRβ. TCRαs on naïve T cells from mice expressing a single TCRβ and various MHCII alleles were sequenced and analyzed for their CDR3α sequences as described in Fig. 3 and 4. CDR3α sequences were defined as the amino acids between and including the conserved cysteine at the C terminal end of the TRAV and the conserved phenyl alanine, tryptophan or leucine in the TRAJ region. Shown are the means and SEMs of 3 independently sequenced identical mice except for H2s DOβWT mice, in which case only 2 mice were analyzed. ***p<0.001, **p<0.01 by one way ANOVA with Newman-Keuls post analysis.

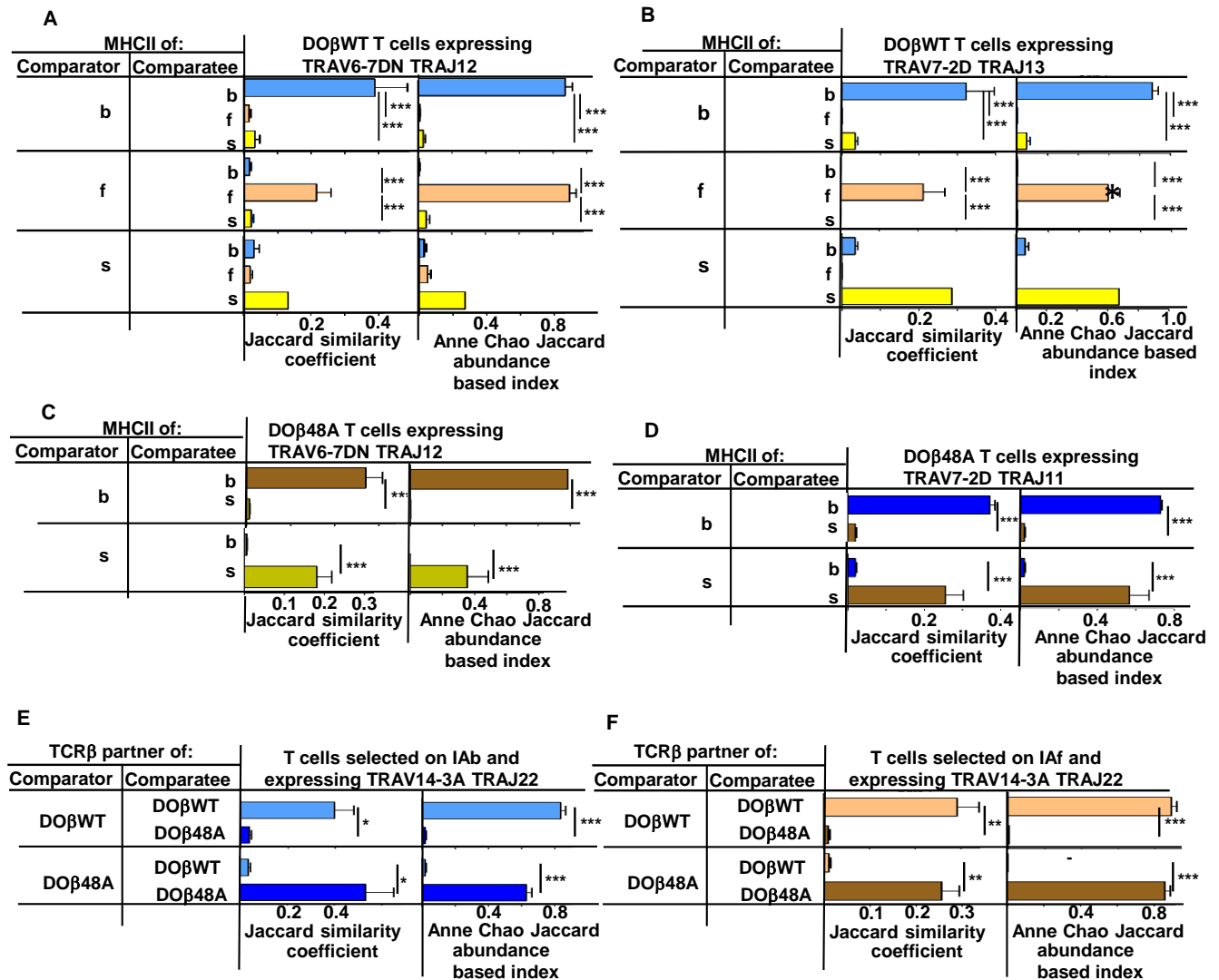**Fig. 8. N region amino acids in CDR3α of naïve CD4 T cells are determined by the selecting MHCII allele and the co-selected TCRβ.** TCRαs on naïve CD4 T cells of mice expressing a single TCRβ and various MHCII alleles were sequenced as described in Figs 3 and 4. Mice were as listed in those Figs. Comparisons were made of N regions derived from the same TRAV TRAJ pair providing that all mice in the comparisons expressed at least 5 different sequences involving the chosen TRAV TRAJ pair. Results shown are the means +/- SEMs of the data from identical mice. Statistical analyses involved one way ANOVA tests with Newman-Keuls post test analyses (A, B) or Student t tests (C-F). *p<0.05, **p<0.01, *** p<0.001.

Supplementary Table 1			
In normal mice, a significant number of TCRα sequences appear on naïve CD4 T cells regardless of the selecting MHCII allele			
% of unique sequences in:	**That are also found in:**		
	B6	**B6.AKR**	**B6.NOD**
B6 (IAb)	35.9+/-4.6	18.9+/-2.9	10.3+/-1.6
B6.AKR (IAk IEk)	35.8+/-2.8		12.2
B6.NOD (IAg7)	36.6+/-2.8	23.1	
% of total sequences in:	**That are also found in:**		
	B6	**B6.AKR**	**B6.NOD**
B6 (IAb)	62.1+/-2.9	44.1+/-2.5	26.7+/-1.6
B6.AKR (IAk IEk)	57.8+/-2.1		27.0
B6.NOD (IAg7)	54.9+/-2.1	41.2	

Naïve CD4 T cells were isolated from the lymph nodes of normal mice of the indicated strains and their TCRα sequences identified as described in the Methods section. Shown are the %s of unique sequences and the %s of total sequences that were shared between pairs of mice of the indicated strains. Data were obtained from 3 independently sequenced B6 mice and one each B6.AKR and B6.NOD animals and are the means and standard errors of the means of the comparisons.
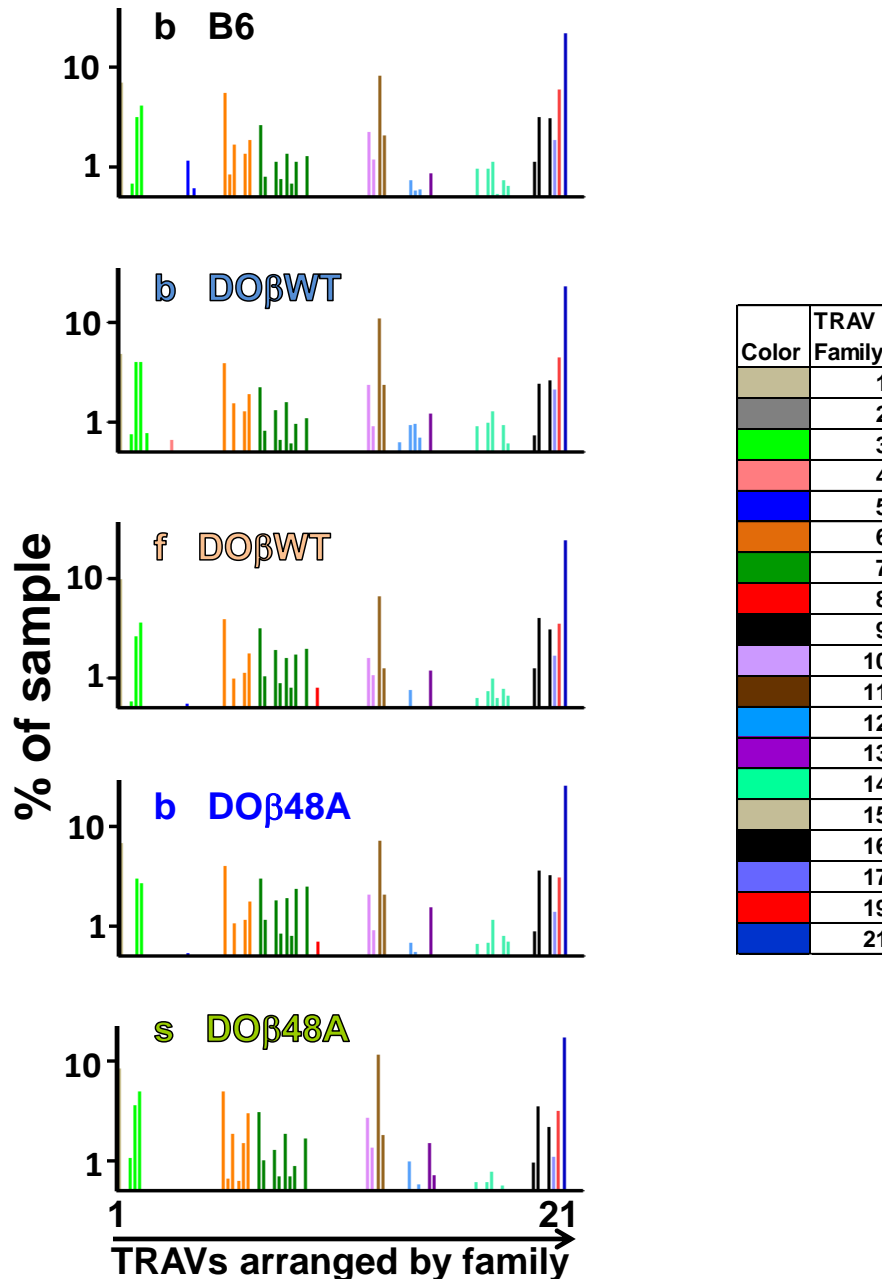
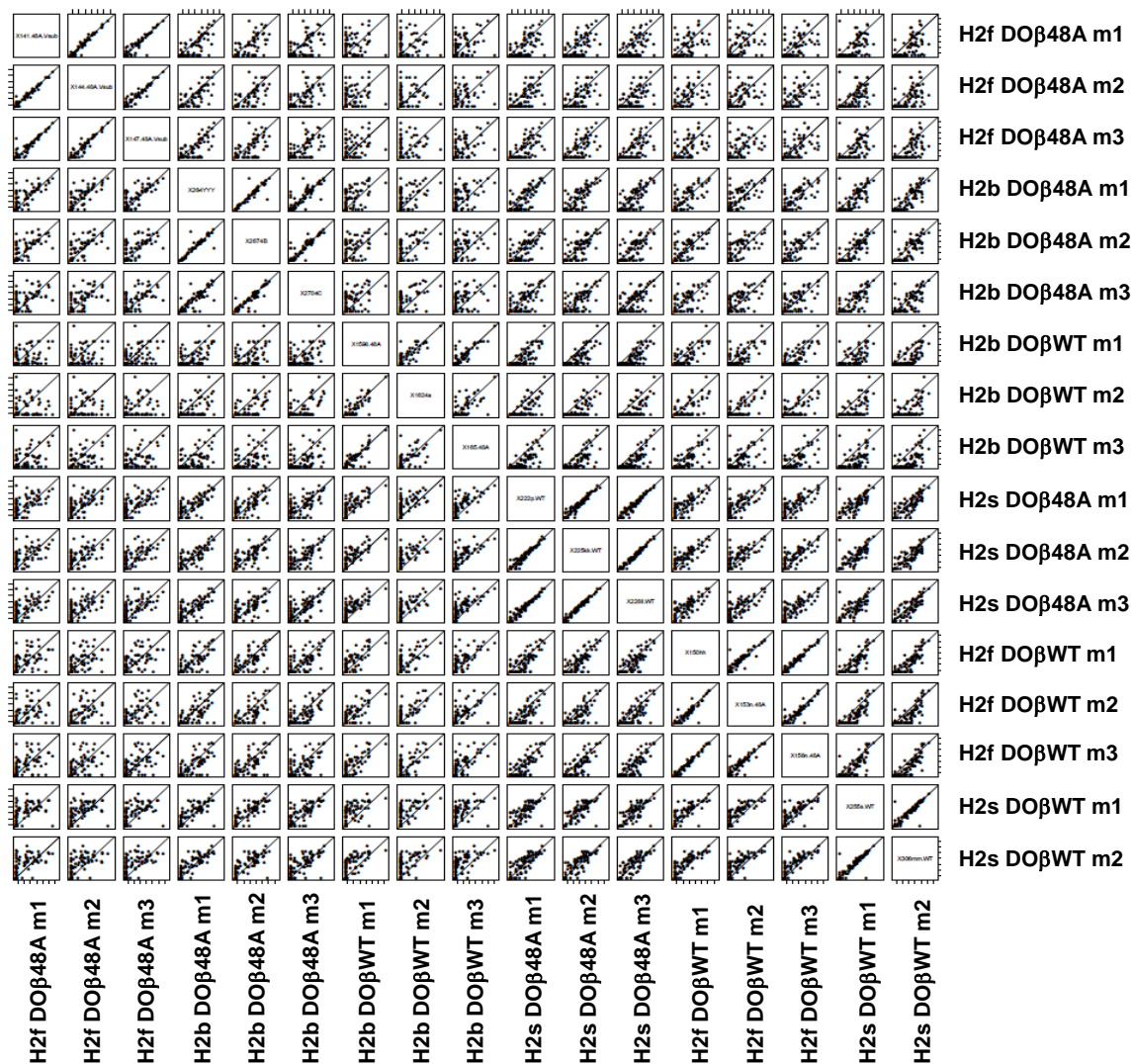| Supplementary Table 2			
Sequences of TCRβ transgenes			
	TRBV (Arden nomenclature)	CDR3	TRBJ
DOβWT	13.2 (Vβ8.2)	CASGSGTTNTEVFF	1.1
DOβ48A	13.2 (Vβ8.2) with Y48 mutated to A	CASGSGTTNTEVFF	1.1
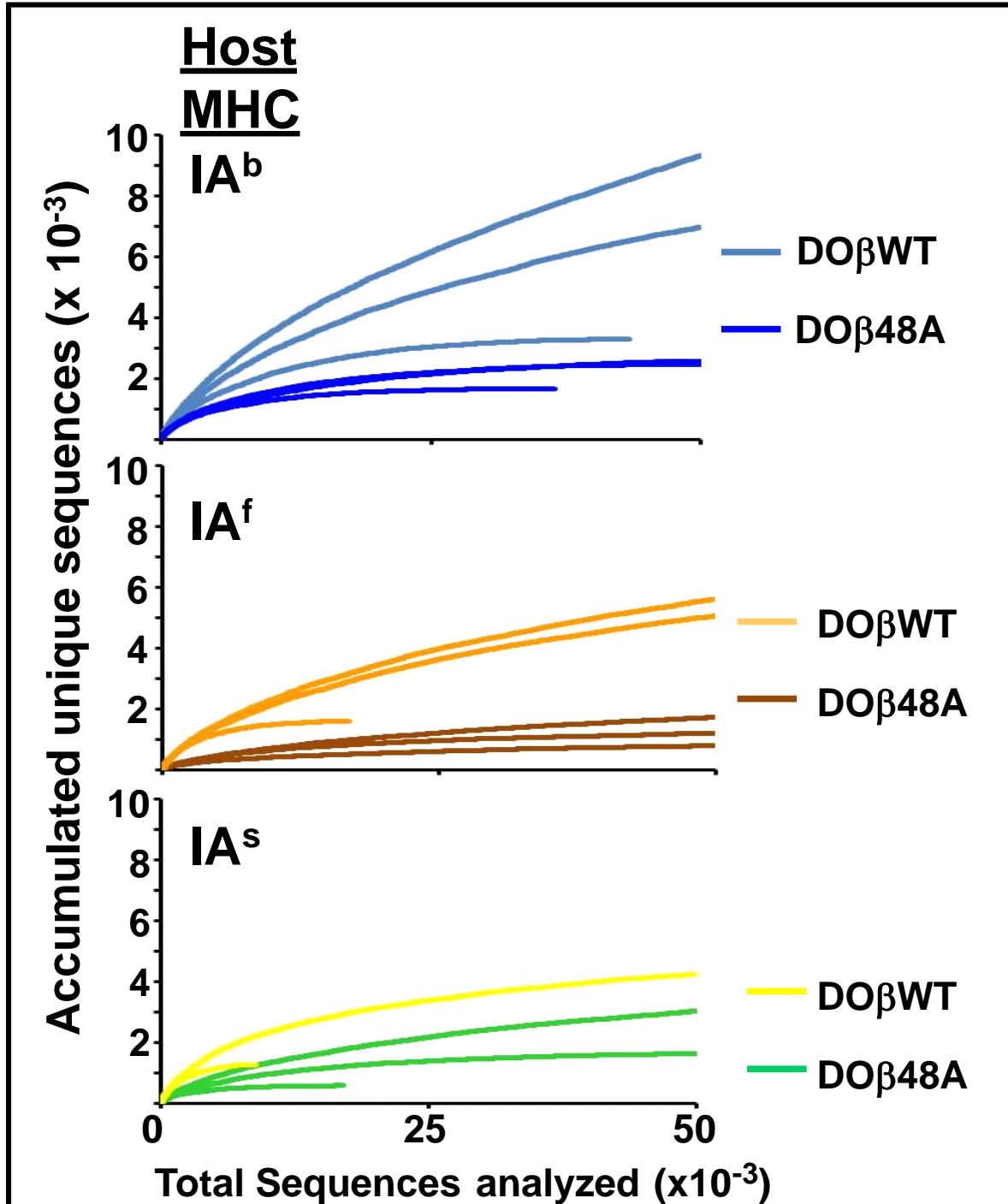
Supplementary Fig. 1 **The naïve CD4 T cells in mice expressing a single TCRα chain express a limited number of TCRα sequences regardless of the MHC allele involved in their selection in the thymus.** Results were calculated as described in Fig. 3. Data shown are for individual mice.
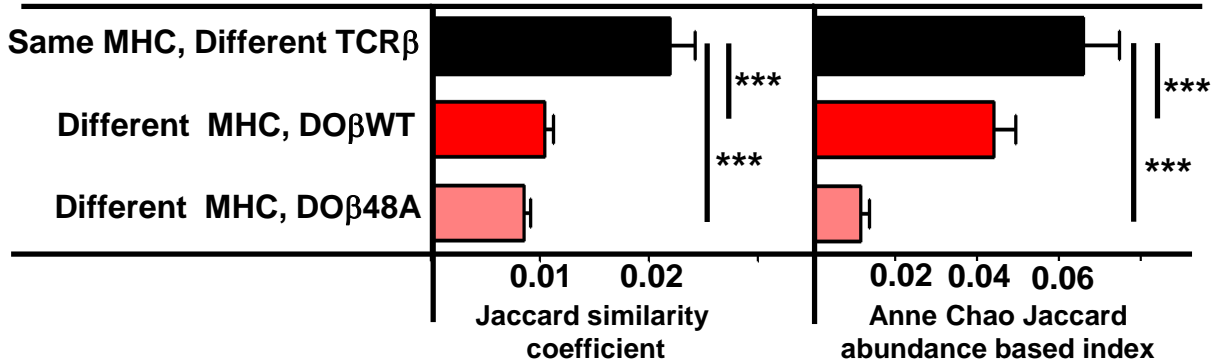
Supplementary Fig. 2 TCRα sequences are somewhat more likely to be shared between T cells selected on the same MHCII allele but differing in TCRβ than between T cells sharing TCRβ but selected on different MHCII alleles. TCRα sequences were obtained and analyzed as described in Figure 4. Data shown are combined as indicated between all the mice illustrated in Fig.4 Results are the means and SEMs of the data. ***p<0.001 by one way ANOVA with Newman Keuls post analysis.

Supplementary Fig 3. Different TRAVs are detected at different frequencies in preselection thymocytes. Preselection DP thymocytes were obtained from individual mice expressing the indicated MHCII alleles and TCRβs. Shown are the %s with which individual TRAVs were detected. Subfamily members of each family are color coded as shown. Note, the TRAVs are ordered by family and not by their position in the TCRα locus.

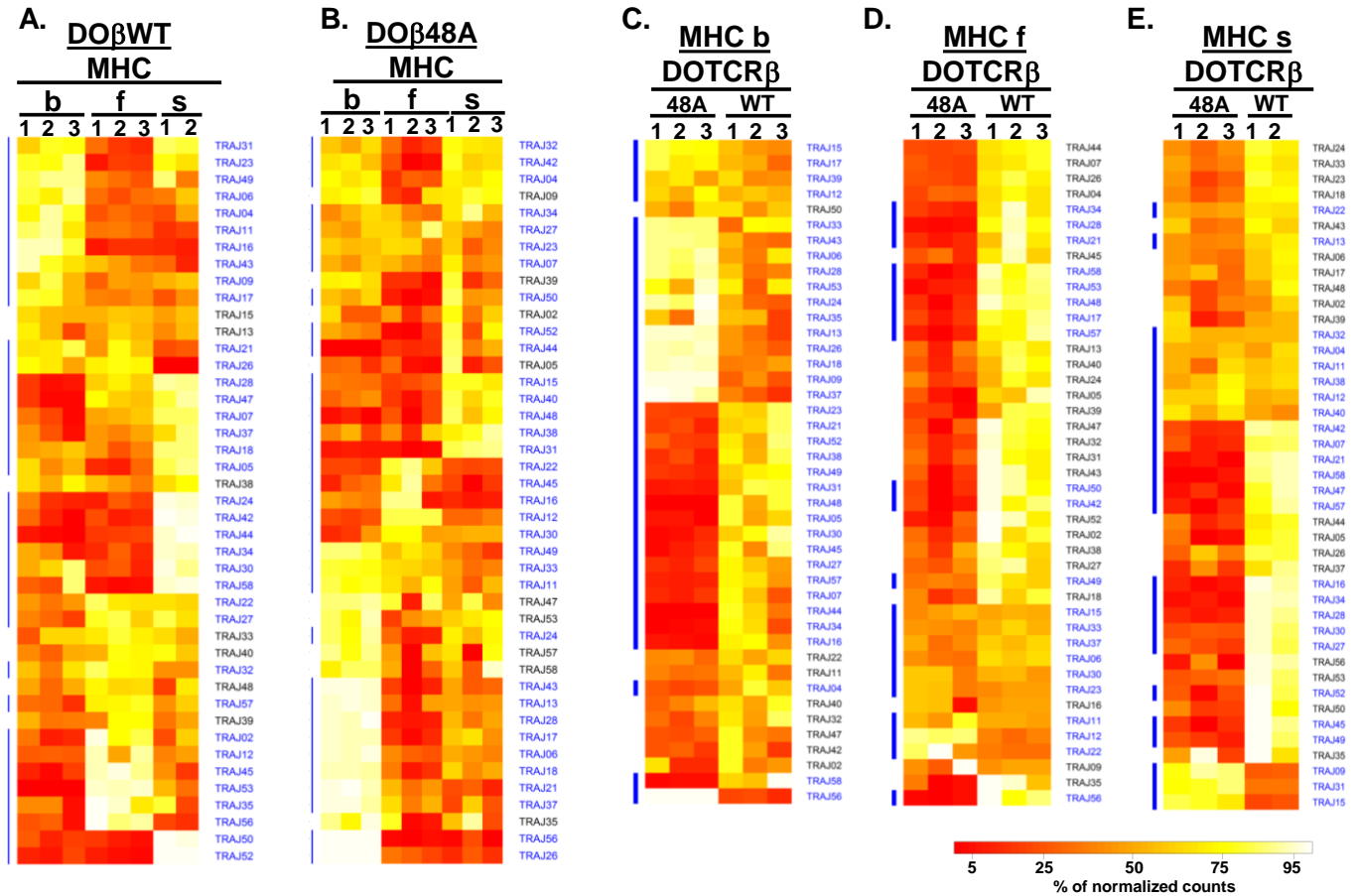**Supplementary Fig. 4. TRAV usage by naive CD4 T cells depends on the selecting MHCII allele and partner TCRβ.** The %s with which individual TRAVs were used by mature naïve CD4 T cells in mice expressing different MHCII alleles were plotted against each other for individual mouse pairs. Individual mice of each genotype were numbered m1-m3, numbers for each mouse are used consistently throughout.
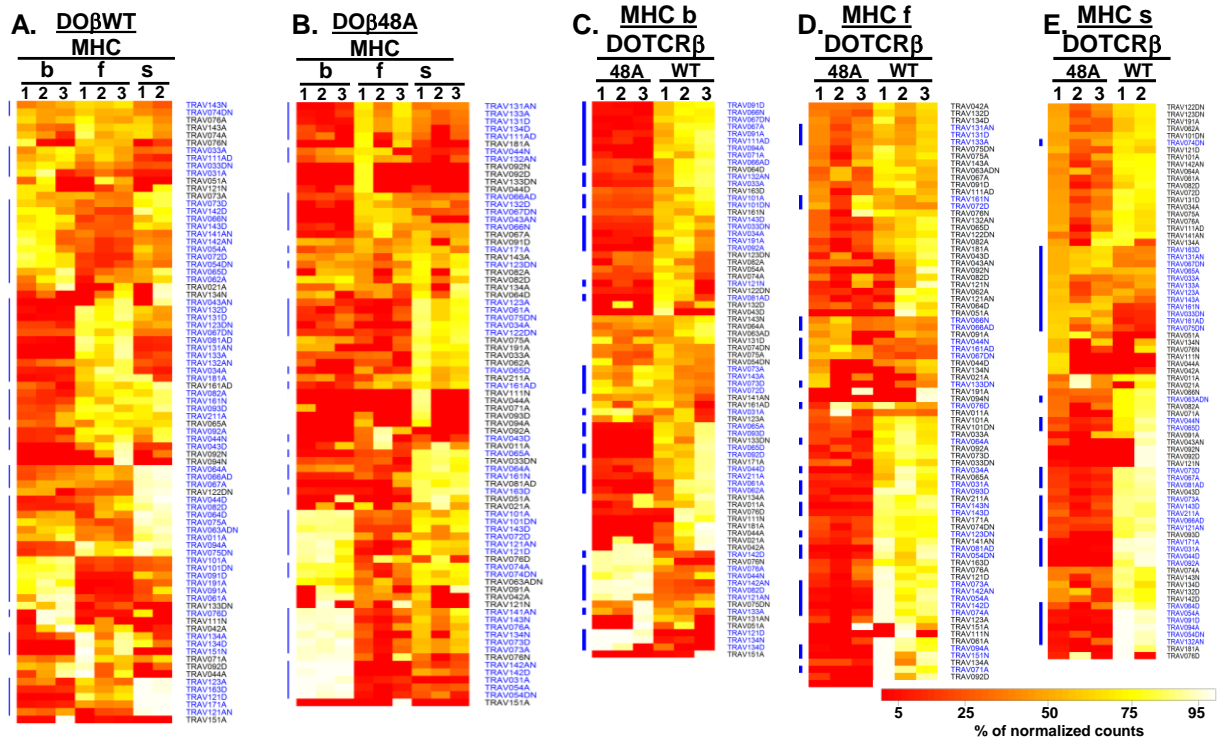
Supplementary Fig. 5 TRAVs are used to different extents by naïve CD4 T cells in mice expressing different MHCII alleles and/or different TCRβs. Data were obtained and analyzed as described in Figs. 3 and 5. DESeq 2 was used to compare the frequencies with which different TRAVs were used by naïve CD4 T cells in mice expressing different MHCII alleles and/or different TCRβs. Differential expression analyses were performed using the DESeq2 package (v1.8.1) in the R language (v3.2.2). DESeq2 fit negative binomial regression models to each feature to compare between groups. Corrections for size factors for each sample (animal) to account for differences in repertoire size were applied. The negative binomial dispersion parameter for each feature was then calculated, sharing information across features with similar expression levels to moderate extreme empirical dispersion estimates. Wald tests were applied to each feature to test for differential expression between groups. Features were considered differentially expressed if they had a Benjamini-Hochberg adjusted p-value (i.e., false discovery rate) <0.05.
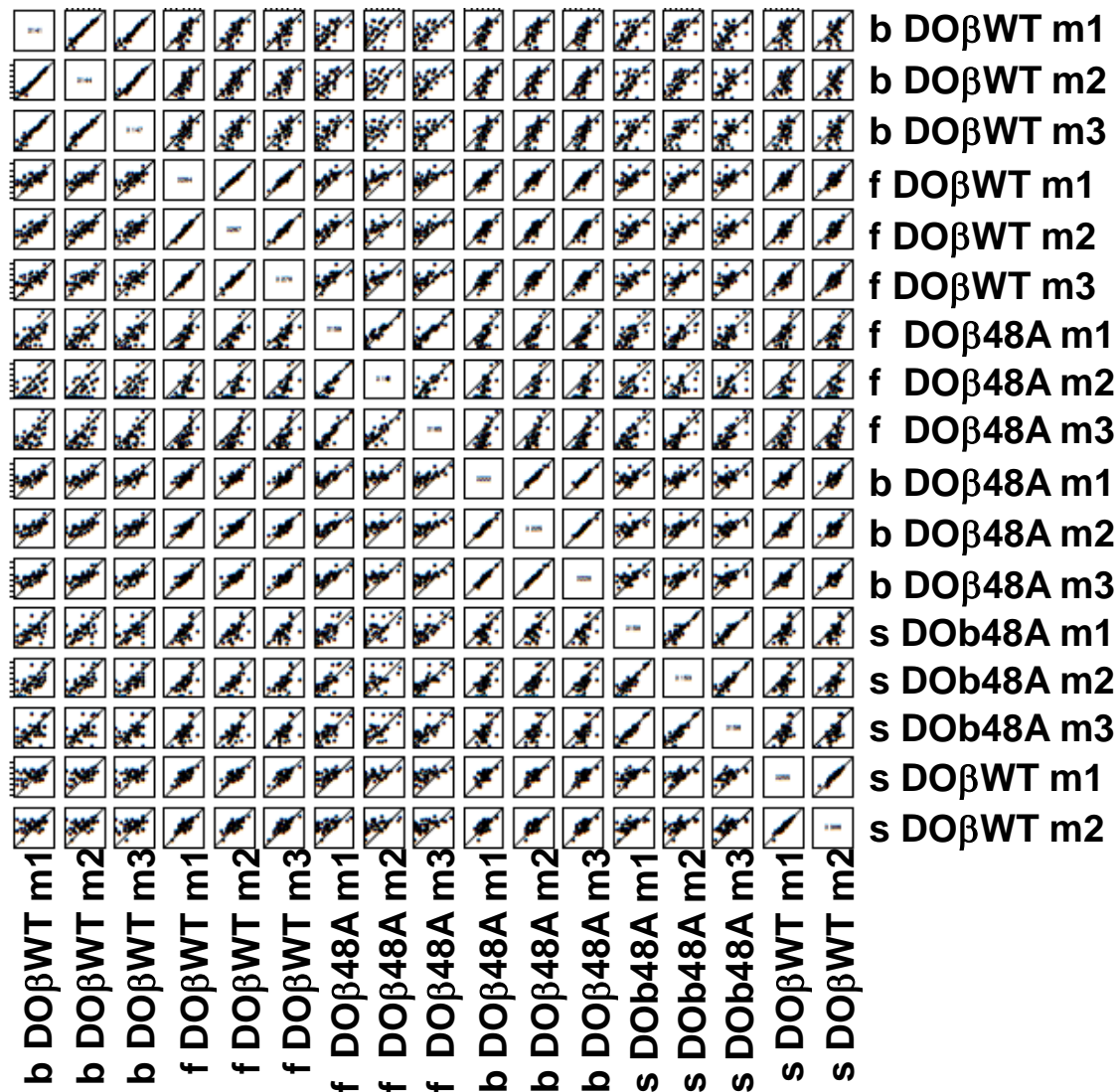
Supplementary Fig. 6. TRAJ usage by naive CD4 T cells depends on the selecting MHCII allele and partner TCRβ. The %s with which individual TRAJs were used by mature naïve CD4 T cells in mice expressing different MHCII alleles were plotted against each other for individual mouse pairs.

Supplementary Fig. 7 TRAJs are used to different extents by naïve CD4 T cells in mice expressing different MHCII alleles and/or different TCRβs. Data were obtained and analyzed as described in Figs. 3 and 5. DESeq 2 was used as described in supplementary Fig. 5 to compare the frequencies with which different TRAJs were used by naïve CD4 T cells in mice expressing different MHCII alleles and/or different TCRβs.
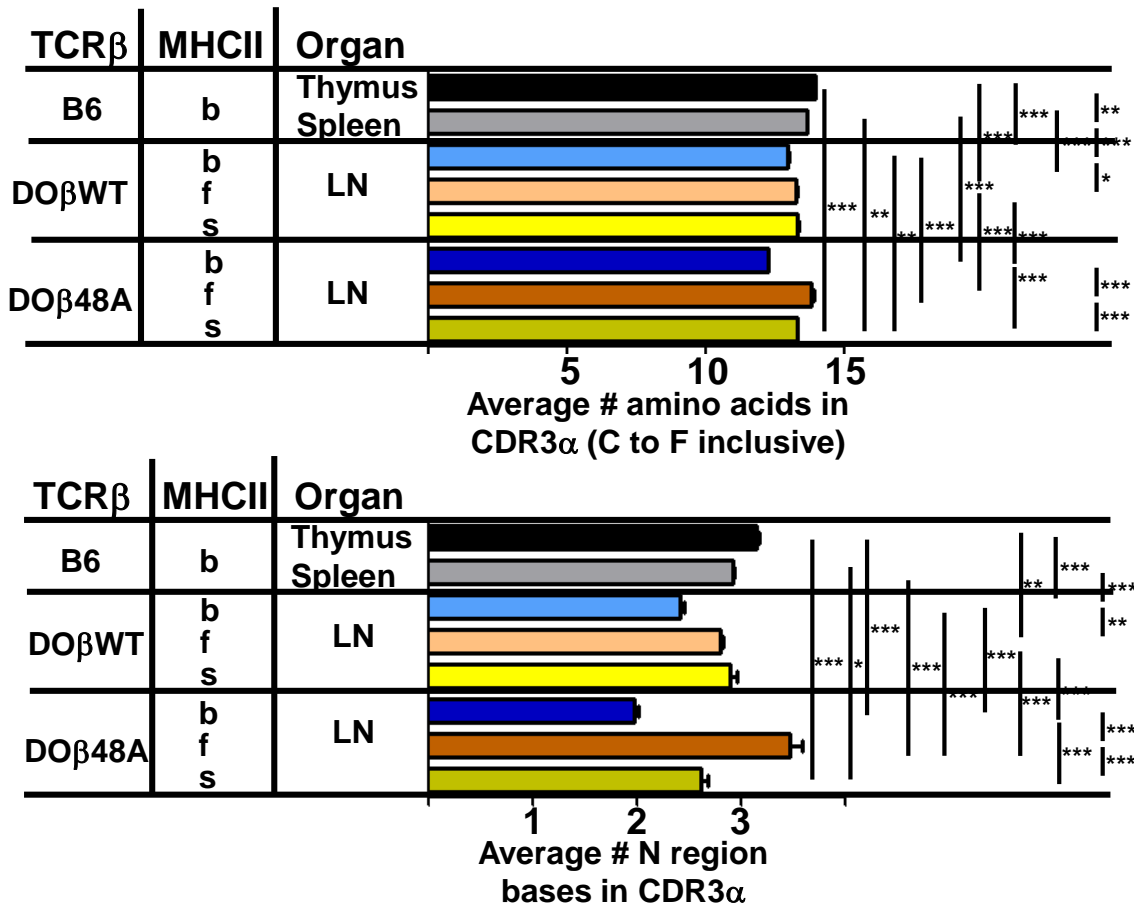
Supplementary Fig 8. CDR3α length on naïve CD4 T cells depends upon the selecting MHC allele and the co-selected TCRβ. Data were obtained as described in Figs. 3 and 4. Shown are the average lengths of CDR3α and N region bases on naïve CD4 T cells coexpressed with the indicated TCRβ. Results are the means and SEMs of 3 mice of the same haplotype except for s DOβWT mice, in which case the results from 2 mice were averaged. Errors are calculated with one way ANOVA with the Bonferroni post test comparing all pairs. * = p<0.05, **=p<0.01, ***=p<0.001. Statistical results are omitted for pairs which were not significantly different from each other.

A

Comparison of TCRα N region sequences on DOβWT T cells expressing the same TRAV and TRAJ

MHCII of:

Comparator	Comparatee		

Comparison of TCRα N region sequences on DOβ48A T cells expressing the same TRAV and TRAJ

MHCII of:

0.1 0.2 0.3
Jaccard similarity coefficient

0.2 0.6 1.0
Anne Chao Jaccard abundance based index

C

TCRβ co-selected partner

Comparison of TCRα N region sequences on T cells expressing DOβWT versus DOβ48A and the same TRAVs and TRAJs

Comparator	Comparatee	IAb	IAf	IAs
DOβWT	DOβWT / DOβ48A			
DOβ48A	DOβWT / DOβ48A			

0.1 0.2 0.3   0.1 0.2 0.3   0.1 0.2 0.3
Jaccard similarity coefficient

D

TCRβ co-selected partner

Comparison of TCRα N region sequences on T cells expressing DOβWT versus DOβ48A and the same TRAVs and TRAJs

Comparator	Comparatee	IAb	IAf	IAs
DOβWT	DOβWT / DOβ48A			
DOβ48A	DOβWT / DOβ48A			

0.4 0.8   0.4 0.8   0.4 0.8
Anne Chao Jaccard Abundance based Index

Supplementary Fig. 9. N region amino acids in CDR3α of naïve CD4 T cells are determined by the selecting MHCII allele and the co-selected TCRβ. TCRαs on naïve CD4 T cells of mice expressing a single TCRβ and various MHCII alleles were sequenced as described in Figs 3 and 4. Mice were as listed in those Figs. The results for samples expressing the same TRAV TRAJ pairs for all comparisons in which each mouse expressed at least 5 different sequences involving the chosen TRAV/TRAJ pair were averaged. Results shown are the means +/- SEMs of the data from identical mice. Statistical analyses for comparisons involving 3 different types of mice involved one way ANOVA tests with Newman-Keuls post test analyses. Student t tests were used to analyze statistically the differences between pairs of mice. ** $p<0.01$, *** $p<0.001$.

Frequency of amino acid use in CDR3αs with lengths (between the C terminal TRAV C and the TRAJ conserved F/W/L inclusive) of:
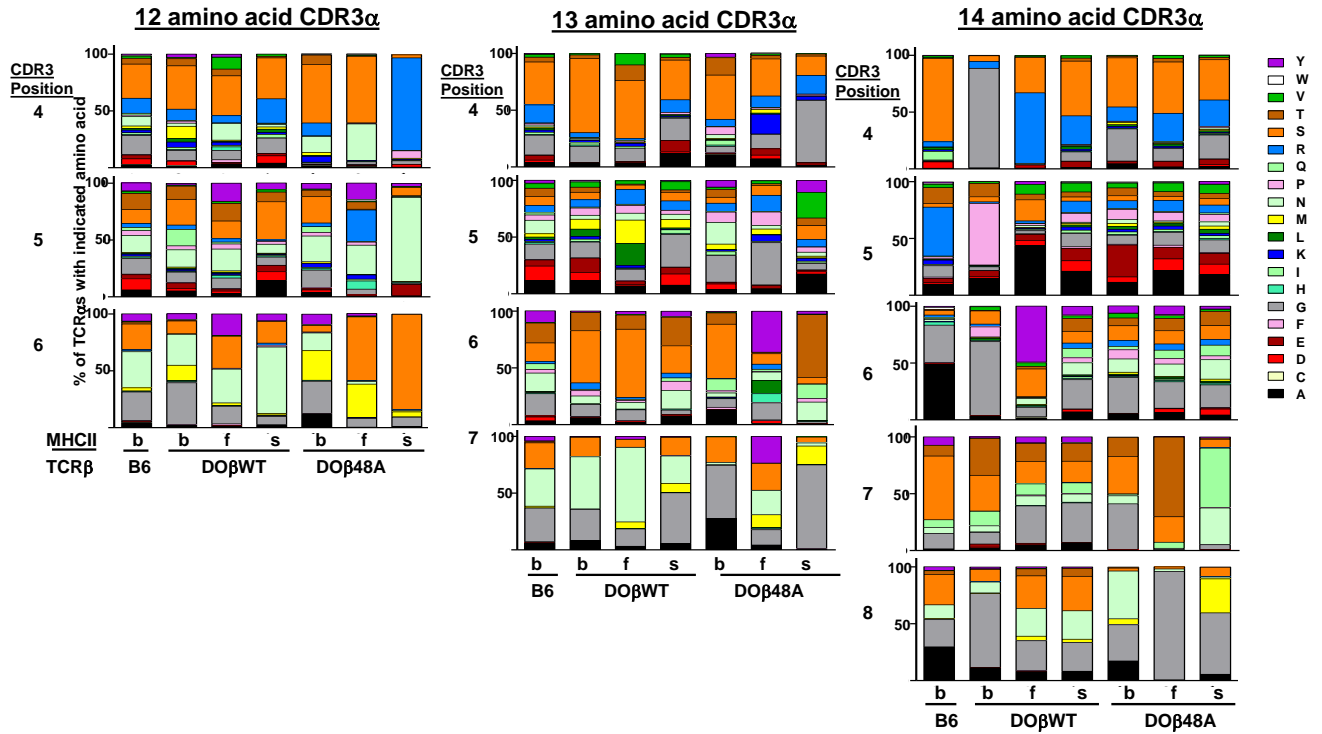
Supplementary Fig. 10 The frequency of amino acid use in CDR3α on naïve CD4 T cells depends on the selecting MHCII allele, the co-selected TCRβ and the length of the CDR3α. Data from TCRα sequences obtained as described in Fig. 4 were used to analyze the frequency with which different amino acids were used at different positions.