THE ROLE OF CONSERVED MHC RESIDUES FOR INTERACTION WITH T CELL ANTIGEN RECEPTORS

by

DANIEL ADAM SILBERMAN

B.S., Rider University, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Immunology Program

2016

This thesis for the Doctor of Philosophy degree by

Daniel Adam Silberman

has been approved for the

Immunology Program

by

Laurent Gapin, Chair

James DeGregori

Jordan Jacobelli

Raul Torres

Andres Vazquez-Torres

John W. Kappler, Advisor

Date: May 20, 2016

Silberman, Daniel Adam (Ph.D., Immunology)

The Role of Conserved MHC Residues for Interaction with T cell Antigen Receptors

Thesis directed by Professor John W. Kappler

ABSTRACT

The interaction of αβT cell antigen receptors (TCRs) with peptides in the context of major histocompatibility complex (MHC) molecules lies at the center of adaptive immunity. Whether the TCRs evolutionary bias guides this specificity, or, rather, processes in the thymus, involving co-receptors and other molecules, select an MHC specific repertoire *de novo* from a random repertoire, is a longstanding immunological question. Here, using nuclease targeted mutagenesis, I address this question *in vivo* by generating three independent lines of knock-in mice with single point mutations of conserved class II MHC amino acids that are often involved in interactions with the germline encoded portions of TCRs. While the TCR repertoire generated in these mutants is similar in size and diversity to wild-type (WT) mice, the evolutionary bias of TCRs for MHC is seen in the shift and preferential usage of some TCR subfamilies over others. Furthermore, the T cells educated on these mutant MHC molecules demonstrate alloreactivity to each other and WT mice, and vice versa, suggesting strong functional differences among these repertoires. Taken together, these results highlight both the flexibility of thymic selection as well as the evolutionary bias of TCRs for MHC.

The form and content of this abstract are approved. I recommend its publication.

Approved: John W. Kappler

ACKNOWLEDGEMENTS

I am forever indebted to John W. Kappler and Philippa Marrack for their incredible mentoring. Without their advice, support, encouragement, and patience, this thesis would not have been possible. Their influence on not only my career, but also my life, will be fondly remembered as long as I live.

I am also grateful to many others at the University of Colorado. My thesis committee of Laurent Gapin, Raul Torres, Andres Vazquez-Torres, James DeGregori, and Jordan Jacobelli has provided valuable advice and support. In particular, Laurent Gapin made me feel like part of his lab for conducting experiments, and he remains my favorite person to discuss science with, be it this thesis or immunology in general. Harsha Krovi and Katie Tuttle in Laurent's lab assisted and energized some of these experiments.

I'd also like to thank my collaborators, Rick Reisdorph, Jen Matsuda, Jimmy Gross (if it wasn't for his expertise, we would never have generated the knock-in mice that are the crux of this thesis), James Crooks, and Janice White. Other students that have enriched my graduate school experience, such as Greg Kirchenbaum, Tony Desbien, Barney Collins, and James Scott-Browne.

My scientific journey started at Bensalem High School under close mentoring from Sam Bates. Upon my arrival at Rider University, I was immediately taken under the wing of immunologist James Riggs, whose contribution to my scientific and personal development is immeasurable.

Finally, I'd like thank my family and friends, whose unconditional love and support have made all this possible. Those that didn't make it to see the end of my educational journey are definitely missed and will never be forgotten.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

INTRODUCTION[1]

Historical Introduction

Immunologists have studied the role of the major histocompatibility complex (MHC) for nearly eight decades, with the earliest studies just hinting at clues, and the later studies finally elucidating the role of the MHC. The first major experiments were conducted by Tyzzer and Little in the early 1900s. They were attempting to discover the rules for transplanting tumors among mice and found that allogeneic tumors are invariably rejected, and the genetics are complex indicating multiple loci (1). Another extremely early study, conducted by Englishmen Peter A. Gorer, regarded a blood group locus that later turned out to be the MHC. These studies used anti-sera to identify 4 blood-group antigens expressed on the surface of red blood cells (2). However, further serological studies of erythrocyte MHC expression revealed that this red blood cell expression has nothing to do with MHC function (3). The erythrocyte MHC expression is actually leftover information remaining in the cytoplasm following expulsion of the nucleus and does not occur in several vertebrate species (4). This less significant feature of the MHC set the table for the study of MHC in transplantation biology and led to the naming of the MHC.

Building on Tyzzer's and Little's experiments, studying tissue compatibility in mice congenic for different histocompatibility reactions, Gorer, Lyman, Snell, and Medawar discovered that cells and tissues carrying one type of MHC were rejected when introduced to an individual whose corresponding molecules were not exactly

[1] Portions of this thesis have been submitted for publication in the *Proceedings of the National Academy of Science*.

the same as those introduced. Snell created congenic strains that isolated individual genes that led to rejection, one of which had the same location as Gorer's RBC antigen II and led to particularly rapid rejection, so he designated the molecules as histocompatibility antigens and gave them a serial number, 2, and thus the genes coding them *histocompatibility 2* (*H-2* as commonly known today) (3). The influence of MHC on the outcome of grafting, however, is only on the fringe of MHC physiological function, as hinted by the earliest signs of immune tolerance, identified by R.D. Owen in bovine twins that were genetically different but shared a placenta and hence their blood systems became mutually tolerant (5). Medawar then studied transplant rejection and showed the same H-2 locus was involved and discussed it in terms of tolerance in F1 mice (6).

The first studies that started to uncover the real role of the MHC orchestrating the adaptive immune response came from the work of Hugh McDevitt and Baruj Benacerraf. They noticed that certain mouse or guinea pig strains, irrespective of the background of the stain, would only mount a response to an individual antigen directly correlating with the H-2 type (7, 8). They named the genes responsible for this as immune response (IR)-governing genes, and mapped them to the MHC. Some suggested that the IR genes encoded the long sought after T cell antigen receptor, but it turned out in fact that these were actually the MHC structural genes themselves (9). To further complicate the understanding of the MHC at this time, there were several reports of complement genes being involved with the MHC locus (10, 11).

At this point, immunologists had had begun to clarify the roles of T and B cells. Claman (12) was the first to show that cells derived from the thymus improved the ability of B cells to make antibodies. Later, in antibody responses to haptens coupled to protein carriers, Mitchinson and his collaborators showed that in order to elicit an antibody response to a hapten, there needed to be simultaneous recognition of a carrier protein by T cells (13). It was also clear at this time that B and T cells recognized different epitopes through immunizations of rabbits and guinea pigs with bovine glucagon in which the antibody response was directed toward the N-terminal end of the protein while lymphocytes divided in response to the C-terminal portion (14). Furthermore, B and T cells cross-reacted differently with different species of red blood cells (15), highlighting the different reactivity of the two populations.

These findings led for the search of the T cell antigen receptor (TCR), but results were confounded by various experiments suggesting T cells secreted a soluble receptor, T cells expressed immunoglobulin molecules, or that the MHC genes were in fact the T cell antigen receptor genes themselves (7). Additionally, experiments that studied the T cells' ability to help B cells' antibody response linked this ability to the H-2 locus (16, 17). In addition, it was shown that macrophages-lymphocyte interactions were also linked to histocompatibility (18).

A clearer picture of the role of the MHC did not come into view until the experiments of Zinkernagel and Doherty (19). Their experiments provided the shift in perspective that was needed to resolve all the apparent contradictions. These experiments showed that while B cells can recognize any antigen, T cells also needed to recognize both antigen and the MHC simultaneously (19). In a

cytotoxicity assay, viral specific T cells would only kill infected target cells if both the T cell and the target were of the same H-2 type.

This leap in understanding left immunologists with an obvious question: was this one receptor on the T cell recognizing both antigen and MHC, or were there two receptors? Some had argued for the two receptor hypothesis (20), but this question was most elegantly answered by the laboratory of John Kappler and Philippa Marrack with a dual TCR expressing T cell hybridoma experiment (21). This experiment showed that fused hybridomas could react with the two combinations of MHC and antigen (MHC A + antigen A or MHC B + antigen B) but not mixed combinations (MHC A + antigen B or MHC B + antigen A). Thus, a single receptor must react with a combination of self-MHC and antigen, which later led to the discovery of the heterodimer TCR (22-24) and the genes (25, 26) of the TCR.

The next obvious question for immunologists to address was how the TCR recognized the combination of self-MHC and antigen simultaneously. There was early evidence that macrophages were somehow processing the antigen for T cells (27). Furthermore, it was shown that this processing event took place early, and cells fixed with paraformaldehyde initially could not process the antigen, but those fixed after 30-60 minutes still could present the antigen (28). Moreover, this processing event could be bypassed by enzymatic digestion of the protein antigen prior to the addition to these presenting cells (29). Ultimately, these studies led to Babbitt, Unanue, Buus, Sette, and Grey demonstrating that these antigenic peptides were specifically binding to the MHC protein linked to the corresponding T cell specificity (30, 31).

Structure of the MHC

From the earliest structures of class I major histocompatibility complex (MHCI) (32, 33) and class II major histocompatibility complex (MHCII) (34) it was evident how these molecules were binding and presenting peptide antigens. The basic structure of an MHC has two membrane-distal domains that form the peptide-binding region (PBR) composed of two anti-parallel α helices resting on a floor of eight β strands, and the two membrane-proximal domains that are immunoglobulin superfamily (IgSF) C1. For MHCI and MHCII, the PBR varies among alleles such to be able to bind a wide variety of peptides in a specialized groove formed by the helices and the β strand floor, but the overall four extracellular domain organization as well as other features such as glycosylation have been conserved evolutionarily among MHC proteins (35). From functional experiments it was known that MHCI and MHCII bound different types of peptides. MHCI was known to bind shorter peptides of 8-9 amino acids in length (36), while MHCII was known to bind a wider variety of longer peptides (30). The structure to function relationship here is obvious. MHCI has a binding groove closed at both ends, such that binds the NH_2 and COOH termini of the peptide are buried and its length therefore, restricted. The MHCII groove is open at both ends allowing peptides of varying length to bind by extending out past the end of the groove.

The amino acid sequence of the PBD determines what peptides each MHC will bind (32-34). Peptides bound in the groove of the MHC place some amino acid side chains facing down into the peptide binding grooves, while others extend upwards out of the groove to interact with the TCR (32-34).

Structure of the T Cell Antigen Receptor

The T cell antigen receptor, highly related to immunoglobulins (Igs) in its structure, is comprised of two chains, α and β, each which are comprised of a variable domain linked to a constant domain (37). Also the genes for the variable domains are created somatically by recombination, similar to those that create the Ig genes. Importantly, it was shown that the combination of these chains into a 96,000 molecular weight glycosylated protein that confers specificity for MHC and antigen (37). The variable α domain gene is formed by recombining an approximately 400 base pair (bp) variable gene segment (Vα, or T cell Receptor Alpha Variable, TRAV) with an approximately 80 bp joining segment (Jα, T cell Receptor Alpha Joining, TRAJ). The recombination is not precise and bases not present in the coding sequence of each segment can be added to the junction (N or P nucleotides). The variable β domain is created similarly, but with three gene segments: a Vβ (T cell Receptor Beta Variable, TRBV), Dβ (TRBD), and Jβ (TRBJ) domain, again with possible N and P nucleotides at the imprecise junctions (38, 39).

In mice, there are about 100 different TRAV and 60 TRAJ segments that allow approximately 6000 different TRAV-TRAJ combinations. Also there are 20 TRBV, 2 TRBD, and 12 TRBJ segments that allow roughly 360 combinations. Therefore pairing random α and β domains leads to 2.2 million gene segment combinations. Taken together with the imprecise recombination, N and P nucleotide additions, the potential repertoire of TCRs is astronomical, with some estimates as high as 10^{12} in mice (35, 38).

The antigen recognition loops of the TCRs and antibodies are called the complementary determining regions (CDRs). In both the TCRα and TCRβ, the CDR1 and CDR2 loops are encoded entirely by the germline V regions. The TCRα CDR3 region is formed from the junction of the TRAV and TRAJ, while the TCRβ CDR3 loop is formed by the junction of the TRBV, TRBD, and TRBJ segments (35, 38, 39). In solved TCR structures, the CDR3 loops are positioned in the center of the face used to interact with peptide-MHC complexes, while the CDR1/2 loops make up the perimeter of the binding surface (35).

TCR Ligand Recognition

When compared to antibodies, TCRs have ~1000-fold lower affinities for their MHC ligands, i.e. ~1-100 μM vs 1-100 nM. The advent of surface plasmon resonance to measure these TCR affinities in the μM range has yielded a lot of information on the thermodynamics of the TCR-peptide-MHC (pMHC) interactions (35). Analysis of the early mutation studies (40, 41) and many more informative TCR-pMHC complex X-ray crystallography structures (39, 42, 43) reveals a generalizable orientation in which the TCRs bind diagonally across the pMHC complex generally at an angle varying between 60 to 90 degrees in relation to the extended peptide in the MHC binding groove. In these structures the Vα and Vβ CDR3 loops are usually centrally located over the peptide, the CDR2 loops over the MHC helices and CDR1 loops positioned to contact the MHC as well as the N-terminal and C-terminal peptide residues, respectively.

The Idea of the Evolutionary Hypothesis

A long standing question in immunology, and the one addressed in this thesis, asks: Are there germline features of TCR V elements that have evolved to favor TCR recognition of MHC ligands? The idea of an evolutionary connection was first suggested by Niels Jerne in 1971. He proposed in a theoretical paper that the lymphocyte receptors had evolved to interact with the alleles of the MHC and that in the thymus cells somatically mutated their receptors to ensure that they no longer interacted with their own alleles but retained the ability to interact with the alleles of others (44). Until the discovery of thymic positive selection, immunologists favored the evolutionary hypothesis due to the overwhelming number of TCRs which recognized some peptide in the context of an MHC (43). Positive selection, the phenomenon by which T cells developing in the thymus must interact with an intermediate affinity with some self-pMHC complex or they die by neglect (45), suggested an alternate hypothesis. Perhaps the initial thymic TCR repertoire was in fact unbiased, but the process of positive selection picked out for development those rare T cells that had acquired this MHC specificity by chance during the somatic recombination process.

Other data appeared that was hard to explain by the evolutionary hypothesis. For instance, in the earliest crystal structures, some thought that there would be a clear way in which all TCRs were biased in their recognition of MHC. However, among these first structures there were no obvious rules by which every TCR was biased, and even some odd TCRs that bound MHC outside the usual diagonal orientation (46).

Furthermore, the genes of the TCR and MHC lie on different chromosomes, so there is no obvious genetic mechanism by which a TCR V gene could maintain co-expression with a specific rapidly evolving MHC gene (43). Furthermore, with the exception of a few TCR V region family members (47-49), most Vα and Vβ elements can be found in TCRs that recognize many different alleles of MHCI, MHCII, or even some non-classical MHC molecules with drastically different stereochemistry (43, 50).

Evidence for the Evolutionary Hypothesis

Yet there is also evidence in favor of the evolutionary hypothesis. Experiments with TCRs that have not undergone positive selection, still find a high bias for MHC ligands. Nonselective maturation of immature thymocytes in the absence of MHC molecules revealed an inherent MHC reactivity in the preselection repertoire that is very high (51). Furthermore, as many as one in five MHC-naive thymocytes show upregulation of activation markers on exposure to MHC-expressing thymic stroma in short-term reaggregate culture (52). More recently, unbiased, high-throughput cloning and retroviral expression of individual pre-selection TCRs provided a direct assessment of TCR selection at the clonal level in vivo, and found that 15% of random TCRs induced signaling and directed positive (7.5%) or negative (7.5%) selection (53).

TCRs also exhibit a phenomenon known alloreactivity, a process by which TCRs ability to recognize peptide-allogeneic-MHC complexes that were not encountered during thymic selection (54). The fact that TCRs crossreact with both self and foreign MHC via alloreactivity with a precursor frequency 100-fold to 1,000-

fold higher than the precursor frequency of any T cell specific for any single foreign-peptide-self-MHC complex suggests a bias of TCRs for MHC. Furthermore, the structural basis of this has been examined and shows that a single TCR recognizes two globally similar but distinct ligands by divergent mechanisms, indicating that receptor-ligand crossreactivity can occur in the absence of molecular mimicry (55). While the molecular basis for alloreactivity remains somewhat enigmatic, an intrinsic affinity of TCRs for the surface of MHC molecules helps to provide a better understanding of how the TCRs and pMHCs interact.

If the TCRs have evolved to recognize MHC, then perhaps there are some clues from the sequences of their CDR regions. Obviously the CDR3 region, which is partially encoded by the genome and partially generated somatically, cannot supply the evolutionarily conserved interaction, so analyzing the CDR1 and CDR2 regions might shed more light (43). While comparison of these V regions in mice and humans does reveal clear homology and transspecies evolution (56), elements which may bias a TCR towards MHC reactivity are not apparent. However, this homology is also in CDR length, which could play a role in biasing the TCR towards MHC. In addition, there is experimental evidence that the CDR1 and CDR2 regions affect MHC restriction. Experiments done on two TRAV9 subfamily members that are biased in their usage in CD4 or CD8 peripheral T cell subsets showed that site-directed mutants of one subfamily member in either the CDR1 or CDR2 were sufficient to change selection from the CD4 subset to the CD8 subset (48). These data suggest that germline-encoded TRAV elements are a major influence on MHC

restriction, most likely by a preferential interaction with either class of MHC molecule.

The most obvious uniform structural feature about how TCRs bind MHCs is the diagonal docking orientation (57). To date, there has only been one natural TCR found to bind an MHC in reverse orientation (58), and only having one example does not fit with the idea that positive selection is all that dictates a TCR's specificity for MHC. However, there is an example in which a docking angle actually limits signaling, although this is not known whether to involve co-receptors or CD3, or perhaps a driving force for an evolution of the TCR bias for MHC (59). Furthermore, the pivot point of the TCR is almost always centered over residues p4-p6 of the peptide, which a random selection event, if still able to signal, would not predict (43).

Just as positive selection plays an important role in the process by which TCRs become MHC specific, negative selection in the thymus also complicates matters. Negative selection is the process by which the thymus eliminates TCRs that react too well with some self-pMHC complex (60, 61). It is possible that in order for a CDR3 to be selected, it must not only lack high affinity for a self-peptide, but also cause steric attenuation of some of the germline encoded interactions of the CDR1 and CDR2 regions. Thus TCRs that have undergone a normal negative selection may not show all the germline encoded interactions structurally because they are being masked by the CDR3 that allowed them to avoid negative selection. It has been shown that mice expressing an MHCII with a single linked peptide have deficient negative selection of $CD4^+$ T cells (62). These mice, however, still express normal self-peptides bound to MHCI so they are not ideal for this experiment, but

examining the TCRs that do get positively selected reveal evolutionary features. Many TCRs from CD4 T cells in this mouse are cross reactive on many different alleles of MHC, a feature one would expect from TCRs bearing stronger germline bias towards MHC molecules (43). Furthermore, the crystal structures of some of these TCRs have been solved and there are interesting germline encoded amino acids that have been studied for their role biasing the TCR towards MHC (63).

Previous TCR-pMHC crystal structures had also been searched extensively for germline interactions that could bias the TCR towards MHC (64, 65). If you change the criteria by which you look for these interactions, some of them start to become apparent. Taking into account that most solved crystals include TCRs that have undergone normal negative selection, fewer of the germline encoded rules might be used for a given structure. Also, the idea that individual V genes will have evolved different mechanisms to recognize both MHCI and MHCII adds more possibilities to conserved interactions but also considers that these interactions may need to be flexible (43). Following these criteria, three key amino acids, TRAV Y29, TRBV Y46, and TRBV Y48, have been identified and studied.

From the structures, careful analysis of how TRAV CDR1 and CDR2 interact with MHC reveals a few conserved features (43). Most striking of these features are TCRs that have TRAV CDR1 that contains a tyrosine at position 29 (Y29). In structures containing either human TRAV9 or mouse TRAV6, this Y29 interacts strongly with βH81 and βT77 of MHCII and is likely to be an evolutionarily conserved feature for biasing TCRs towards MHC. Among other TRAVs, there are other common amino acids, such as Y31, which also very often interact with the α helices

of either MHCI or MHCII. In the TRAV CDR2, a serine at position 51 is another commonly conserved feature of many Vαs and it also very often interacts with MHC (43). Overall, inspection of the crystal structures just starts to scratch the surface of understanding the germline encoded interactions of the TRAV elements towards MHC.

On the TCR Vβ side, similar analysis of the crystal structures has led to the best evidence to date for germline encoded biasing of TCRs towards MHC. These data stem from the fact that mTRBV13 family members are very often used in mice, and their orthologous human elements are also very prevalent. A feature these family members have in common is the presence of Y46, Y48, and E54 in the CDR2 region. Because of the large number of crystal structures containing these amino acids, rules for how TRBV13 family members recognize different MHC molecules have been illustrated (64). Perhaps the most interesting of these amino acids is Y48, and it is structurally seen to contact the MHC in similar manners in numerous TCR-pMHC interactions (43).

While the structures give hints as to what rules govern how specific TCR V regions are biased towards MHC, functional experiments were needed. The first series of experiments to functionally test the role of these amino acids in thymic selection came from the laboratory of John Kappler and Pippa Marrack (66). They showed that retrogenic T cells with a fixed TRBV13 TCRβ chain expressing mutations of any of these three amino acids in the TRBV family had impaired selection of $CD4^+$ T cells in the thymus, and most prominently, the βY48A had the largest reduction of $CD4^+$ T cells and also a reduction of $CD8^+$ cells. Furthermore,

when they combined a TCRα chain from a TCR with the βY48A mutation and then added back the WT βY48 TCRβ, it became auto reactive against I-A^b, the molecule it was selected on (66). These experiments clearly show that in this system, these germline encoded amino acids on the TCR are critical for their recognition and specificity for MHC, yet some questions remain. These cells express a fixed βCDR3 that had already undergone thymic selection, and these amino acids are not present on every family of TCR Vβ regions, so the generality of the interactions have been questioned.

Some Evidence against the Evolutionary Hypothesis

The strongest opposition to the evolutionary hypothesis has come from the laboratory of Al Singer. The hypothesis his lab attempted to test involved the role of the co-receptors, CD4 and CD8, for generating TCRs specific for my MHC. They argued that because CD4 and CD8 sequester the kinase Lck, required for TCR signaling, that the co-receptors themselves, with specificities only for MHC ligands, might impart the specificity of the TCR for MHC (67). To test this, they created a mouse called the quad-deficient mouse, which lacks co-receptors and MHC molecules. In this system T cells develop that are specific for non-MHC ligands (67). While these experiments prove that TCRs are able to recognize non-MHC ligands, a fact that is not surprising given that their CDR3 regions can generate so much diversity, they do not directly address whether the TCR is biased towards recognition of the MHC. However, they went on to show that one of these TCRs that contains a VβY48 is actually specific for the cell surface molecule CD155 and mutation of Y48A abrogates that recognition (68). Most recently, they added back mutant Lck that

does not associate with co-receptors, and shown that the T cells that get selected are greatly auto-reactive, even against non-MHC, and also against MHC ligands (69).

Other experiments also argue that the role co-receptors contribute to ligand recognition trumps an intrinsic ligand bias. The Dyson lab created another system in which they mutated the germline TCR complementarity determining regions via recombination cassettes *in vivo*. They showed that engagement with MHC ligands during thymocyte selection and peripheral T-cell activation imposes remarkably little constraint over TCR structure. They argue that such versatility is more consistent with an opportunist, rather than a predetermined, mode of interface formation (70).

In addition to these functional experiments in complicated systems, there has been some structural evidence hinting at the adaptability of TCRs (71, 72). These experiments attempt to show structurally how TCRs are not biased towards MHC, although it is not clear that their conclusions are confirmed by their own data (73) or necessarily mutually exclusive from the evolutionary hypothesis.

Rationale for My Experiments

The function of the *Mhc* in immune responsiveness is also reflected in its genetic polymorphism. Polymorphism is the presence at any given time of a larger than expected number of genetic variants in a population (35). In humans, *HLA* genes exhibit an extraordinarily high degree of polymorphism which is thought to have evolved because of the battle between pathogens and the immune system (74). Looking evolutionarily at vertebrates, MHC genes have the highest level of polymorphism compared to other genes by both number and diversity of alleles (75).

This degree of polymorphism makes sense as a heterozygous pool of antigen-presenting elements in a given individual might allow the binding and presentation of antigenic peptides derived from a wide variety of environmental pathogens. Limited polymorphism would make the entire population susceptible to a chance infectious agent against which all individuals would be unable to respond, whereas widespread polymorphism would be expected to allow the antigen-presenting cells of at least a proportion of the population bind and present antigens derived from invading pathogens effectively (35). It turns out in fact that there is an optimal amount of MHC diversity, at least in stickleback fish, which is most likely attempting to balance the benefits of peptide presenting diversity with the ability to overcome negative selection (76). Examining the location of the polymorphisms fits well with the idea that the polymorphisms exist to extend the range of peptides, since the amino acids that vary widely in MHC are those most often found in the peptide binding grooves (77). The solvent-exposed residues on the α helices of the MHC, on the other hand, become interesting candidate to study because of their relative lack of polymorphism, and also their conservation between species (78, 79).

Looking at these solvent-exposed residues on the α helix of the MHC in structures reveals that they are often contacted by the TCR, both germline encoded and somatically rearranged amino acids in all of the CDR loops (42, 43). Functional studies mutating some of these residues have previously been shown to affect specific mature T cell responses to both MHCI (71) and MHCII (80). The extent of the effect, however, was dependent on the mutation chosen, the specific TCR, and even the peptide the MHC was presenting.

In the current study I assessed the importance of several MHCII conserved docking sites for TCRs by introducing specific point mutations into mouse I-A^b MHCII α or β genes both *in vitro* and *in vivo*. *In vitro* these mutations had little effect on the collection of self-peptides bound by the mutant I-A^b, but often disrupted MHC plus peptide recognition by T cells specific for a variety of foreign or self-peptides. *In vivo,* mice carrying these MHC point mutations developed TCR repertoires that were similar in size to those of WT mice, but with altered TRAV or TRBV gene usage. Furthermore, in vitro in mixed lymphocyte reactions, T cells from each of the WT and mutant mice responded strongly to antigen presenting cells from the other mice, but not to their own cells. I discuss these results in relation to the current ideas about the role of evolution vs. somatic selection in framing the T cell repertoire.

CHAPTER II

MATERIALS AND METHODS

Mutant MHC I-A^b α and β Chains

Plasmids encoding MHCII I-A^b α and β chains were available in the Kappler/Marrack lab. MHC mutations were cloned by overlapping primers using engineered restriction sites. The I-A^b α chain was cloned into a murine stem cell virus (MSCV)-based retroviral plasmid with an internal ribosome entry site (IRSE) plus Thy1.1 as a reporter. The I-A^b β chain was cloned into a similar MSCV vector with a GFP reporter. These MSCV vectors were also available in the Kappler/Marrack lab (66, 81, 82).

Retroviral Packaging

Retroviral plasmids were co-transfected into Phoenix cells with pCLEco accessory plasmid using Lipofectamine 2000 (Invitrogen, Carlsbad, California) using manufacturer's instructions. Retrovirus-containing supernatants were collected at 48-72 hours after transfection and filtered through a .45 micron filter to remove cell debris (83).

MHC Expressing Cell Lines

MHC constructs were expressed by retroviral transduction of an antigen presenting cell line, M12.C3. M12.C3 cells are derived from a BALB/C B cell lymphoma that was selected for loss of I-A expression (84) although they contain a functional I-A^d α chain. For retroviral infection of M12.C3, 105 cells were spin-infected with retroviral supernatants containing 8 µg/ml of polybrene for 90 min at 37

degrees Celsius. Cells were expanded in culture, subsequently cloned by limiting dilution, and clones of equal MHC expression were chosen.

Bulk T Cell Hybridomas

Antigen specific T cell hybridomas were generated by immunizing mice with the desired antigen emulsified in complete Freund's adjuvant. The para-aortic lymph node cells were isolated seven days later, expanded in culture for three days with the same antigen the mice were immunized, and cultured in IL-2 for five days. After this *in vitro* culture, activated T cells were fused to BWα$^-$β$^-$, a variant of the fusion partner BW5147 generated to lack both TCRα and TCRβ chains (85).

Hybridoma Stimulation

For stimulations, 5 x 10^4-1 x 10^5 hybridomas were cultured with different stimuli for 4-24 hours in 200 µl in 96 well microtiter plates. Hybridoma responses were measured by several different readouts. Primarily an IL-2 ELISA was done using antibody clone JES6-1A12 to capture (eBioscience, San Diego, CA) to capture and biotinylated antibody clone JES6-5A4 (eBioscience) with streptavidin conjugated to horseradish peroxidase (Jackson ImmunoResearch, West Grove, PA) to detect presence of the biotinylated secondary. CD69 activation was also measured on the bulk hybridomas by flow cytometry using clone H1.2F3.

Mass Spectrometry

WT and mutant I-A^b proteins were immunoprecipitated from lysates of roughly 10^9 of the transduced M12.C3 cells using antibody clone Y3P. Peptides were eluted in 2.5M acetic acid and separated from beads, Ab, and class II MHC molecules by

passage through a 10,000-day cutoff ultrafiltration unit (Millipore, Bedford, MA) and subjected to MS or MS-MS analysis as previously reported (98).

Untargeted Differential Analysis

Samples were analyzed by LC/MS in duplicate. Data were mined in Mass Hunter Qualitative Analysis software (Agilent Technologies) using an untargeted feature finding algorithm. Extracted molecular features, e.g. isotopes, charge states, and adducts, were deconvoluted and assigned to peptide masses. For each peptide, all assigned charge states were combined and the peak areas for each were summed and the total area (volume) was used for relative peptide abundance. Extracted peptide molecular features were imported into Mass Profiler Professional (Agilent Technologies). Peptides were aligned by mass and retention time, and filtered for minimum relative frequency across replicates. Subsequent analysis was limited to peptides detected in all wild type replicates.

LCMS

Peptides were analyzed via LC/MS/MS or LC/MS on an Agilent QTOF (model 6520) mass spectrometer with an HPLC-chip interface. All HPLC components were Agilent 1100. Buffer A of the nano pump was comprised of HPLC grade water and 0.1% formic acid, buffer B was 90% acetonitrile, 10% HPLC water, and 0.1% formic acid. The loading pump utilized 3% acetonitrile, 97% HPLC grade water and 0.1% formic acid. A so-called “long” HPLC chip (Agilent G4240-62006) was used which consisted of a 40 nL enrichment column and a 150 mm x 75 um analytical column combined on a single chip.

In order to ensure that comparable amounts of peptide material were analyzed for each fraction, initially 0.5 µl of each sample was analyzed to establish overall relative abundance. Initial runs were performed with a 10 min LC/MS method (gradient was from 3-30% B buffer over 0-5 minutes). Areas of extracted TICs were used to calculate injection volumes for each sample. Subsequent runs were with 'injection volume-corrected' samples and used a 30 min LC/MS/MS or LC/MS/ method (gradient from 3-30% B buffer over 0-25 minutes).

Raw LC/MS/MS data were extracted and searched using the Agilent Spectrum Mill search engine, and spectra were searched against a Swissprot mouse database. "Peak picking" was performed within SpectrumMill with the following parameters: signal-to-noise was set at 10:1, variable modifications were searched for oxidized methionine and deamidated asparagine, maximize charge state for peptides was set at 7, precursor mass tolerance of 20 PPM and product mass tolerance of 50 PPM. Matched peptides were filtered with a score > 6 and a Scored Peak Intensity of > 60%.

Flow Cytometry

Cells, either *ex vivo* or hybridomas, were pre-incubated with anti-CD16/CD32 antibody producing hybridoma supernatant (clone 2.4G2). Cells were stained under saturating conditions with antibodies to mouse TCRβ (clone H57-597), CD4 (clone GK1.5), CD8 (clone 53-6.7), CD25 (clone PC61), CD44 (clone IM7), CD5 (clone 53-7.3), CD69 (clone H1.2F3), CD24 (clone M1/69), B220 (clone RA3-6B2), CD11b (clone M1/70), γδ TCR (clone GL3), CD62L (clone MEL-14), Vβ8.x (clone F23.1), Vβ8.2 (clone F23.2), Vβ8.3 (clone 1B3.3), Vβ8.1/2 (clone MR5-2), and Vα2 (B20.1)

purchased from eBioscience or BD Pharmingen (San Diego, CA) or generated in house. Cells were analyzed by flow cytometry on FACScan, LSR II, or LSRFortessa (BD Biosciences, San Diego, CA).

Generation of Knock-in MHC Mutant Mice

Female mice were superovulated using pregnant mares serum gonadotropin (PMSG) and human chorionic gonadotropin (HCG) (Calbiochem, San Diego, CA) to mimic the follicle-stimulating hormone (FSH) and luteinizing hormone (LH) respectively. Superovulated females were placed in a cage containing a stud male. The following morning, females were checked for the presence of a vaginal plug. Embryos from plugged females were isolated and cumulus cells were digested and washed in hyaluronidase to produce clean single cell embryos for microinjection (86).

Pronuclear injection of the single cell embryo was done at the Mouse Genetic Core Facility (MGCF) at National Jewish Health in collaboration with Jennifer Matsuda and James Gross. Pronuclei were injected with ZFN mRNA (Sigma-Aldrich, St. Louis, MO) and an oligo homology directed repair template (Integrated DNA Technologies, Coralville, IA) or a ds-DNA fragment generated and purified in house as follows. A plasmid encoding roughly 2000 bp of homology plus the desired mutation was digested with restriction enzymes (New England Biolabs, Ipswich, MA) engineered to remove the bacterial portion of the plasmid. Digestion products were run on an ultra-pure seaplaque agarose gel (Lonza, Basel, Switzerland). The correct DNA band was excised and purified sequentially using a Zymoclean Gel

DNA Recovery kit (Zymo Research, Irvine, CA), PureLink PCR Purification Kit (Invitrogen, Carlsbad, CA), Millipore Dot Dialysis (Millipore,Billerica, MA).

Injected embryos were surgically implanted into pseudo-pregnant females. Following birth, pups were screened for presence of the desired mutation by PCR, restriction enzyme digestion, and sequencing.

Sequencing of TCR Repertoire

Naïve CD4 T cells were stained as described above and sorted at the National Jewish Health Flow Cytometry Core Facility on the Synergy (Beckman Coulter, Pasadena, CA). RNA was isolated using the RNeasy kit (Invitrogen, Carlsbad, CA). cDNA was made using the SuperScript VILO kit (Invitrogen, Carlsbad, CA). For sequencing of TCRα chains, a two-step PCR was done which added the machine oligonucleotides as well as barcodes for the sequencing runs. The first PCR included a reverse oligonucleotide in the constant region of the α chain, and a mixture of forward oligonucleotides that together cover all the different TCRα family members. The sequence or sequences for each TRAV family is as follows:

TRAV01, GAGGGAACCTTTGCTCGGGTC;
TRAV02, TATGAAGGGCAAGAAGTGAAC;
TRAV03, CArGTCTTCAGTTGCTTATGA;
TRAV04, TGCTCTGAGATGCAATTTTwC;
TRAV05a, GGTGGAACAGCTCCCTTCCTC;
TRAV05b, ATGGCTGCAGCTGGATGGGA;
TRAV06a, GGACAAGGTCCACAGCTCCT;
TRAV06b, GGAGAAGGTCCACAGCTCCTC;
TRAV06c, GTCCAATATCCTGGAGAAGG;
TRAV07, AGCAGAGCCCAGAATCCCTCA;
TRAV08, AAAGAGCCAATGGGGAGAAG;
TRAV08, GAATAGTCAACTAGCAGAAG;
TRAV09, AGCTGAGATGCAAsTATTCCT;

TRAV10, ACTTACACAGATACTGCyTCA;
TRAV11, CACAGGCAAAGGTCTTGTGTC;
TRAV12, GCTGAACTGCACCTATCAGA;
TRAV13, TGGTTCTGCAGGAGGGGGArA;
TRAV14, GTCCCCAATCTCTGACAGTCT;
TRAV15, ACTGTTCATATrAGACAAGT;
TRAV16, TGGAGAAGACAACGGTGACA;
TRAV17, GTTATTCATACAGTGCAGCAC;
TRAV18, ACCGCACGCTGCAGCTCCTCA;
TRAV19, TACCCTGACAACAGCCCCACA;
TRAV21, GTAGCCACGCCACAATCAGTG.

The second PCR included just one forward oligonucleotide to add the machine oligonucleotide and a reverse oligonucleotide to add the barcode. For TCRβ chains, one PCR that contained a forward oligonucleotide priming all three TRBV13 family members at an identical region as well the sequencing machine oligonucleotide (CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATGCTGAGGCTGATCCATTA), as well as a reverse oligonucleotide that primed the Cβ region and contained the machine oligonucleotide and the barcode (CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATCTTGGGTGGAGTCACATTTCTC).

Processing of TCR Sequences

In house software was developed for processing the raw sequences generated in our sequencing runs. First TCR sequences were identified and short sequences were excluded. Next, the family, subfamily, TRAJ (or TRBJ) usage was identified if possible, and sequences that could not be fully identified were excluded. Finally, the CDR3 region was identified and out of frame sequences or those with a termination codon were excluded. Upon further processing, known sequencing errors were identified in the CDR3 regions and corrected if three nucleotides past

the error completely match the assigned V or J gene, the mismatched nucleotide is changed to the germline encoded one.

Statistical Analyses

Differential expression analyses were performed using the DESeq2 package (v1.8.1) (87) in the R language (v3.2.2) (88). This widely-used package was designed for RNASeq experiments, but its statistical model can be appropriate for count data generated by other high-throughput methods. For example, DESeq2 is used to test for differential abundance of microbial DNA sequences (89), and its predecessor DESeq has been used for differential methylation analyses (90, 91). In conjunction with the DiffBind package(92), DESeq2 has also been used for differential binding analyses of ChIP-Seq data (93).

DESeq2 fits negative binomial regression models to each feature (each TCR family or subfamily in the present context) to compare between groups (genotypes). First, it calculates size factors for each sample (animal) to account for differences in repertoire size. Then it estimates the negative binomial dispersion parameter for each feature, sharing information across features with similar expression levels to moderate extreme empirical dispersion estimates. Finally, with the computed size factors and dispersions it performs Wald tests on each feature to test for differential expression between groups (genotypes). Features were considered differentially expressed if they had a Benjamini-Hochberg (94) adjusted p-value (i.e., false discovery rate) <0.05.

Two- and three-dimensional principal components plots were created using the pca3d (95) package in R. Because the results of a principal components

analysis using raw counts would be overly impacted by the highest-expression features, the raw counts were first regularized using the rlog transformation in DESeq2. This transformation is similar to a log2 transformation but returns finite values when counts equal 0. The principal components of this regularized data were computed by the prcomp function, which first normalizes the features to have mean 0 and variance 1.

Other statistical analyses were conducted using GraphPad Prism version 6.0 or Microsoft Excel. All statistics shown were generated using Unpaired T-tests unless otherwise specified.

CHAPTER III

SELECTING I-A^b MUTATIONS

Introduction

In this chapter, I will show the data from which I chose the mutations that were most interesting to test in thymic selection. To accomplish this, I will explore candidate mutations on I-A^b molecules. Then they will be expressed in a system that tests their effects on mature T cell responses. Ultimately, I will chose mutations that have a large effect on these mature T cell responses and also fit with my idea that they affect these responses through germline encoded interactions with TCRs. I will then proceed with these mutations of interest into the next chapter, in which these mutations are introduced into mice.

Selecting Candidate I-A^b Mutations

Although MHC genes are very polymorphic, most of the changes in amino acid sequence occur in the peptide binding grooves of the MHC proteins. For any given MHC protein there are few changes, often even between species, in the amino acids that are frequently engaged by TCRs (43). These residues are near monomorphic in the mouse I-A molecules found in the majority of laboratory strains. Ten of these amino acids are highlighted in the I-A sequences shown in Table 1 and their positions on the α1 and β1 domains of I-A^b shown in Figure 1. Thus these amino acids may have been conserved during evolution for TCR interaction.

To begin to study the relative importance of these amino acids in TCR recognition of peptide-I-A^b complexes, I mutated separately each of these 10 residues. Non-alanine amino acids were replaced with alanine (A) while alanines

Table 1: Alignment of I-A Haplotype Sequences

Haplotype	Aα chain					Aβ chain				
	55	57	61	64	68	66	70	73	77	81
Consensus IA Sequence	D	P	Q	G	G	L	Q	N	I	A	T	G	K	H	-	Y	L	E	R	T	R	A	E	L	D	T	V	C	R	H
IAb	•	•	•	•	•	•	•	•	•	V	V	•	•		E	I	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IAd	E	•	•	•	•	•	•	•	•	A	E	•	•		E	I	•	•	•	•	•	•	V	•	•	A	•	•	•	•
IAk	E	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IAz	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	Y
IAu	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	Y
IAr	•	•	•	•	•	•	•	•	•	V	V	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IAf	•	•	•	•	•	•	•	E	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IAs	•	•	•	•	•	•	•	•	•	•	•	•	Y		•	•	•	•	Q	•	•	•	•	•	•	•	•	•	•	•
IAq	•	•	•	•	•	•	•	•	•	•	•	•	•		E	I	•	•	•	•	•	•	V	•	•	•	•	•	•	•
IAg7	•	•	•	•	•	•	•	•	•	A	E	•	•		•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•

An alignment of I-A haplotype sequences of residues on the helices. Amino acids that are the same as the consensus sequence are represented with a dot. The solvent-exposed residues potentially interacting with TCRs are highlighted, cyan for I-$A^b\alpha$ and magenta for I-$A^b\beta$, showing their lack of polymorphism .

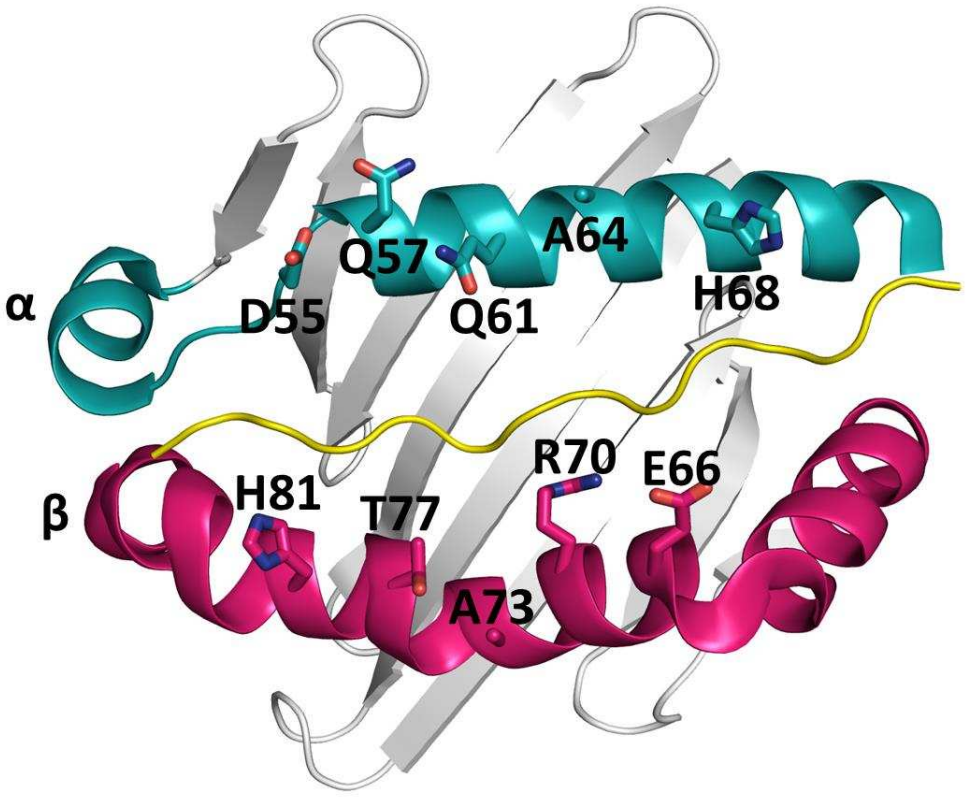

Figure 1: Position of I-A^b Residues That Interact with TCRs
The residues in both I-A^{b}α and I-A^{b}β chains that interact with TCRs are highlighted. This crystal structure (PDB: 3C5Z) specifically depicts the side chains for the residues that were mutated in this study. (63)

were replaced with glutamine (Q). Genes encoding either the I-A^b α or β chain bearing one of the mutations, paired with the corresponding WT I-A^b α or β gene, were transduced into an MHCII deficient B cell lymphoma, M12.C3 (84, 96) to create antigen presenting cells (APCs) expressing the mutant I-A^b molecules. M12.C3 cells transduced with both of the WT I-A^b genes as a positive control and with the WT I-A^b α gene only were used as a negative control. Additionally, since the M12.C3 cells carry a functional I-A^d α chain from the original M12 BALB/c lymphoma that could pair with any introduced I-A β chain, I also prepared M12.C3 transduced with the WT I-A^b β chain only to control for the possible activity of the I-$A^{d/b}$ mixed molecule.

All of the M12.C3 transductants were cloned at limiting dilution and surface expression of I-A^b was confirmed by flow cytometry. Because each mutation might have affected differently the epitopes recognized by individual mAb, I stained the cells using a variety of anti-I-A^b specific mAbs. Figure 2 shows the data for the 17/227 mAb, the antibody least affected by the mutations. With this mAb, the WT and mutant I-A^b cells all stained with a mean fluorescence intensity (MFI) 10-30 fold higher than that of the negative controls. The MFI for the cells expressing the I-$A^{b/d}$ mixed molecule was much lower. Thus all the mutants were expressed and able to be screened in an antigen presentation assay.

Screening I-A^b Mutants

To first test the effect of the mutations, I screened the panel against individual hybridomas and titrated their peptides. My results, similar to what has been observed before with mutations of these or similar residues on different MHC molecules, showed that a wide variety of these mutations affected the response, but

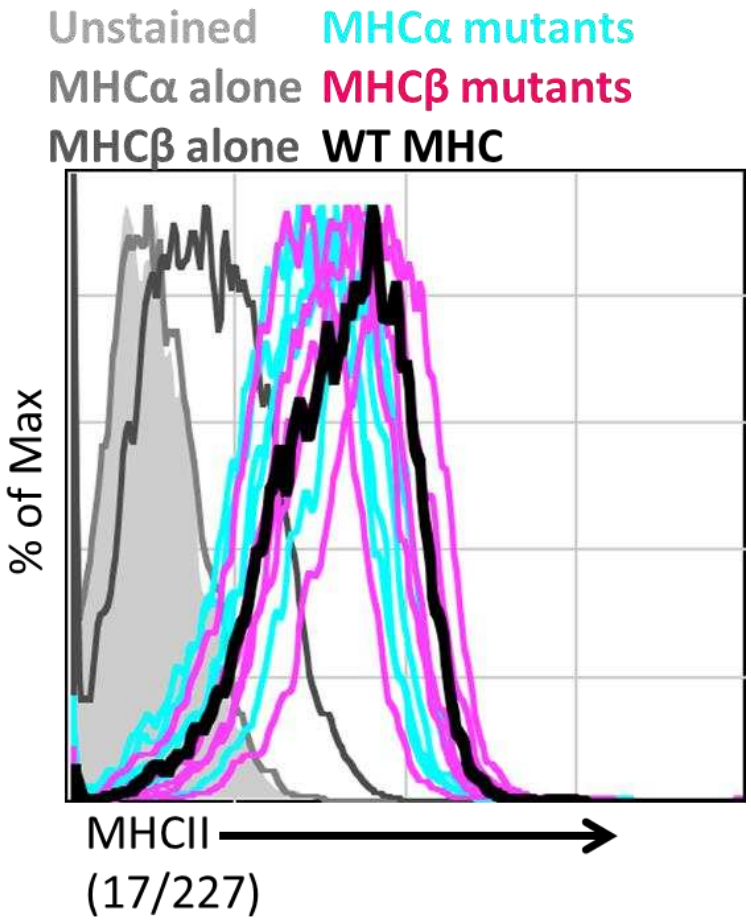

Figure 2: Expression of Mutant MHC Panel in M12.C3
M12.C3 cells were transduced with either WT or mutant I-A^b chains along with WT chains coding for the paired chain. The staining portrayed here was conducted using an antibody clone (17/227) that recognizes I-A^b and is not sensitive to any of the mutations tested. This staining is representative of at least three independent experiments conducted using M12.C3 cells expressing the different mutant I-A^b chains.

that different hybridomas were sensitive to different mutations. One of these hybridomas, B3K506, which recognizes the 3K peptide, derived from the Eα peptide (3K, FEAQKANKAVD) is shown responding to 3K presented by all the MHC mutants (Figure 3). This pattern of mutations affecting individual T cell responses is a representative of other hybridomas as well, but due to the variation of sensitivity to different mutants among hybridomas, I abandoned this approach. I sought a "gestalt" for which mutations would be mostly likely to affect selection, so I decided to devise a system in which the responses of many different T cells could be assessed simultaneously.

The generation of hybridomas was done in collaboration with Janice White of the Kappler/Marrack Lab. C57BL/6 mice were immunized separately with one of five different antigens (Table 2). Seven days later, T cells from the draining lymph nodes of the immunized mice were restimulated with their cognate antigens and expanded *in vitro*. Sets of bulk T cell hybridomas, each specific for I-A^b and one of the immunizing antigens were prepared by fusion of the expanded T cell blasts to the $\alpha\beta^-$ BW5147 thymoma cell line. The preparations were named for their target MHC-II allele, I-A^b, and antigen, listed in Table 2. The BB5 hybridoma was made against a peptide that contains one epitope of the B5 protein of vaccinia virus (B5, AA FTCDQGYHSSDPNAV). The BNP fusion was made against a longer peptide of the nucleoprotein of Lymphocytic Choriomeningitis Virus (LCMV, NP 309-326, AA SGEGWPYIACRTSIVGRA) and contains more than one epitope that can be presented by I-A^b. The BHEL and BKLH fusions were made against whole proteins containing lots of epitopes from hen-egg lysozyme (HEL) and keyhole limpet

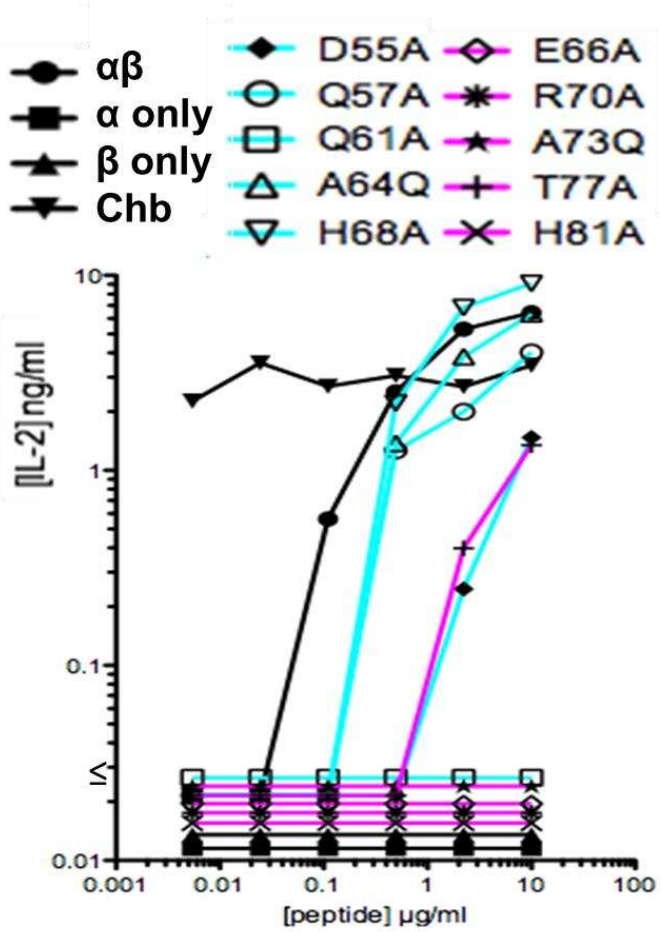

Figure 3: Response of B3K506 is Reduced by Many of the MHC Mutations
B3K506 hybridoma cells were incubated with M12.C3 cells expressing the different I-A^b MHC mutations. These hybridoma cells are specific for 3K peptide presented by WT I-A^b. Chb-2.4.4 cells expressing WT I-A^b were used as a positive control in addition to the M12.C3 cells transduced with WT I-A^b. This experiment was performed once but is representative of how other I-A^b-restricted hybridoma cells respond, although the mutations to which each individual hybridoma was most sensitive were different. IL-2 production by the hybridomas was used as a readout of activation. The black lines represent WT I-A^b chains, the cyan lines depict I-A^bα mutations, and the magenta lines portray I-A^bβ mutations.

Table 2: Summary of Bulk Hybridomas

Bulk hybridoma	Protein or peptide antigen	Sequence
BB5	Vaccinia B5 peptide	FTCDQGYHSSDPNAV
BNP	LCMV NP peptide	SGEGWPYIACRTSIVGRA
BHEL	Hen egg lysozyme	whole protein
BKLH	Keyhole limpet hemocyanin	whole protein
BMOG	Myelin oligodendrocyte glycoprotein peptide	MEVGWYRSPFSRWHLYRNGK

List of the bulk hybridomas used in these experiments. Peptides used for immunization are listed with their sequence as well as whole protein immunogens. All mice used were H-2^b, as denoted by the first letter of each hybridomas name.

hemocyanin (KLH), respectively. Finally, the BMOG fusion was made against a myelin oligodendrocyte glycoprotein (MOG 35-55, AA MEVGWYRSPFSRWHLYRNGK) large self-peptide implicated in experimental autoimmune encephalomyelitis (EAE).

Initially, my goal was to measure activation on an individual cell level, such that I could stain with Vα or Vβ antibodies and identify specifically what was most affected by the mutation. To measure activation, I first looked at CD69 upregulation at different time points following stimulation. I realized quickly that on an individual cell level there was non-TCR mediated CD69 upregulation as almost the entire population of the bulk fusion would shift in response to their given antigen. To address this background a Vβ17 hybridoma was added as a negative control because that gene is not functional in B6 mice, and this hybridoma can be detected with antibody clone KJ23 to distinguish it from the bulk population. Upregulation of CD69 on it was observed, albeit to a lesser extent than the hybrids in the bulk fusion. Thus it was not the proper control for the bystander activation. Because of this I tested intracellular IL-2 expression by staining as well as IL-2 concentration in the supernatant, both of which yielded some information, but ultimately I decided just to use CD69 MFI as a measure of total activation in the cultures (Figure 4).

The bulk T cell hybridoma preparations responded well to the immunizing antigen presented by the WT I-A^b APCs. On average, about 50 fold more of the T cell hybridomas in the bulk populations responded to their immunizing antigen plus M12.C3 bearing WTαβ I-A^b than to control M12.C3 with only I-A^b WTβ. Nearly all of the responses of the peptide or HEL-specific bulk T cell hybridomas were

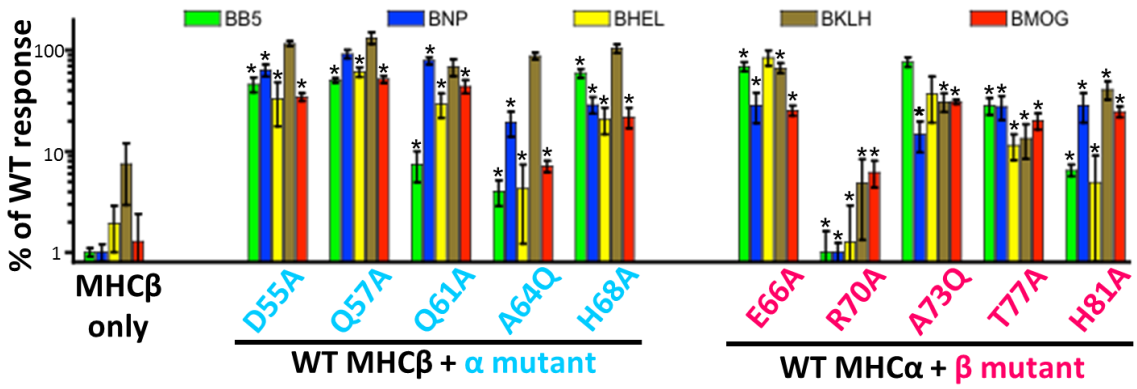

Figure 4: Summary of Bulk Hybridoma Stimulation by Mutant MHC
Bulk T cell hybridomas were stimulated with their cognate antigens presented by WT or mutant I-A^b-expressing APCs. CD69 upregulation was used as a measurement for hybridoma activation. The mean fluorescence intensity (MFI) of CD69 on the hybridomas after stimulation in various conditions is plotted in the graph after normalization to WT responses. Data are representative of 3-4 biological replicates per group and statistical significance is indicated by an * for $p < .05$ by a one sample t test with true value 100.

significantly reduced when using the mutant APCs instead of the WT APCs, again with their immunizing antigen. The responses by bulk KLH-specific T cell hybridomas were also reduced 1.5 to 16 fold depending on the β chain mutant, but not the α chain mutant APCs. This might be explained by the fact that KLH is a very large protein with many potential I-A^b binding epitopes and therefore less sensitive as a group to any one mutation. Consistent with this relative lack of sensitivity to α chain mutants, some of the bulk KLH-specific cells were also cross-reactive to KLH presented by APCs bearing the mixed I-A molecule in which the I-A^d α chain replaced I-A^b α chain.

These results confirmed and extended our lab's previous mutational studies (80) since they showed that the conserved amino acids on the MHCII helices that I had mutated were not required for MHCII surface expression. However, in agreement with previous work, they are often important for TCR docking during a peripheral $CD4^+$ T cell response leaving the possibility open that their conservation might be because they part of germline encoded favorable MHCII docking sites for TCRs. From these results I concluded that T cells that have been selected in the presence of these amino acids do not respond as well to antigens when presented by MHCII without these amino acids.

Choosing the Mutations to Test on T cell Development

I wanted to select mutations from this group for *in vivo* studies to see if they also had an effect on T cell thymic development. I considered the four mutations that most consistently inhibited T cell activation: A64Q on the α chain and R70A, T77A and H81A on the β chain. α64A is invariant in mouse and human MHCII molecules

and creates a docking "cup" for TCR Vβs that contain a tyrosine (Y) at position 48 of CDR2 (43, 63). βT77 and βH81 are adjacent on the I-A^b β chain α helix (Figure 1A). βT77 is invariant in common mouse I-A and I-E alleles and in human HLA-DR and HLA-DQ alleles. In TCR/MHC structures, these two amino acids are often contacted by the TRAV CDR1 loop (43). βH81 is highly conserved in MHCII molecules examined on all mammalian species, but other data implicate βH81 in the activity of mouse H-2DM and human HLA-DM, the proteins that catalyze endosomal peptide loading into MHC (97) and in TCR/MHC structures often makes a surface exposed H-bond to the peptide backbone. Therefore, I decided not to mutate this amino acid in my future experiments. βR70 is nearly monomorphic in all mice I-A alleles (Table 1), but not conserved in mice I-E alleles or in the MHCII alleles of other species. In nearly all published TCR/I-A structures it lies in the central region of the TCR footprint interacting with the TCR CDR3s and therefore, might be expected to influence somatic CDR3 selection during thymic selection, but perhaps not so much germline Vα and Vβ usage. Accordingly, I choose αA64Q as βT77A the primary mutations to test my hypothesis and βR70A and as a potential control.

Effects of the βT77A, βR70A and αA64Q Mutations on the I-A^b Peptides Bound to I-A^b

Before proceeding to *in vivo* experiments with these mutants, I had to consider the possibility that, despite the predicted lack of a direct role for these I-A^b amino acids in peptide binding, they might indirectly change the spectrum of I-A^b presented self-peptides. Since thymic selection involves MHCII occupied by a collection of self-peptides, changes in the spectrum of peptides bound to MHCII

would confound the studies described here, that are intended to examine the effects of MHCII mutations on direct interactions between TCRs and the MHC protein only. To determine to what extent the mutations I created altered the peptides bound to the MHCII proteins, I compared the repertoire of peptides bound to WT I-A^b vs. the other three I-A^b mutants expressed in the M12.C3 transfectants.

As described in the Materials and Methods, the WT and mutant I-A^b proteins were immunoprecipitated from lysates of the transduced M12.C3 cells. Peptides were eluted from these preparations and subjected to MS or MS-MS analysis as previously reported (98). The mass spectrometry conducted in this thesis was in collaboration with Rick Reisdorph and the National Jewish Health Mass Spectrometry Core. A preliminary MS-MS analysis of the peptides isolated from WT and mutant I-A^b established that the method isolated peptides with an I-A^b binding motif (Table 3), many of which had been found in previously in similar experiments (99) (100). This finding served to validate the method of peptide isolation, and suggest that the I-A^b mutations did not affect the I-A^b peptide binding motif.

To compare the peptides bound to WT versus mutant I-A^b, immunoprecipitations and elutions for each sample were performed and analyzed with duplicate runs by MS. Limited MS-MS was also performed to again confirm the presence of the I-A^b binding motif in the peptides. A list of the peptides with identical HPLC retention times and calculated masses that were present in three separate WT I-A^b samples was compared to those in duplicate runs of mutant samples (Table 4). Nearly all of the total peptide intensities found in the WT I-A^b samples were also identified in all of the mutant I-A^b samples. To determine if unique peptides were

Table 3: List of Peptides Identified by MS/MS

Protein	Sequence
	-1 1 2 3 4 5 6 7 8 9
IAb beta chain	(G)RPDAEYWNSQPEILER(T)
IAb beta chain	(G)RPDAEYWNSQPEIL(E)
IAb beta chain	(R)HNYEGPETHTSLRR(L)
IAb beta chain	(R)HNYEGPETHTSLR(R)
IAb beta chain	(R)HNYEGPETHTSL(R)
IAb beta chain	(H)NYEGPETHTSLR(R)
IEd alpha chain	(F)ASFEAQGALANIAVDKA(N)
IEd alpha chain	(S)FEAQGALANIAVDK(A)
IEd alpha chain	(S)FEAQGALANIAVDKA(N)
IEd alpha chain	(F)ASFEAQGALANIAVDK(A)
Lrp10(lipid internalization)	(R)APETRSQVTPSVPSEALDD(S)
Lrp10(lipid internalization)	(R)APETRSQVTPSVPSEALDDS(T)
Lrp10(lipid internalization)	(R)APETRSQVTPSVPSEALDDST(G)
Invariant Chain	(A)KPVSQMRmATPLLMRPM(S)
Invariant Chain	(A)KPVSQMRMATPLLMRP(M)
Invariant Chain	(A)KPVSQMRMATPLLMRPM(S)
Invariant Chain	(A)KPVSQMRMATPLLMRPMS(M)
SLAMf7 (lymphocyte adhesion)	(Q)IPKVVKSPSSLPAKP(L)
SLAMf7 (lymphocyte adhesion)	(Q)IPKVVKSPSSLPAKPLVP(R)
SLAMf7 (lymphocyte adhesion)	(Q)IPKVVKSPSSLPAKPL(V)
Ly86 (associated with CD180/TLR4 at surface)	(G)EQIYYAGPVNNPGLD(V)
Ly86 (associated with CD180/TLR4 at surface)	(G)EQIYYAGPVNNPGLDVPQ(G)
E3 ubiquitin-protein ligase MGRN1	(Q)FPYVTPAPHEPVK(T)
E3 ubiquitin-protein ligase MGRN1	(Q)FPYVTPAPHEP(V)
E3 ubiquitin-protein ligase MGRN1	(Q)FPYVTPAPHEPVKT(L)
Gamma-enolase (neuron cell survival)	(Y)TAKGLFRAAVPSGASTG(I)
Gamma-enolase (neuron cell survival)	(T)AKGLFRAAVPSGAST(G)
Germinal center B-cell-expressed transcript 2	(F)SSHAFQQPRPLTTPY(E)
V-ATPase subunit S1 (acidifying compartments)	(R)LPYTASSGLMAPRE(V)
Sortilin-related receptor (endocytic receptor)	(W)GIDPYDQPNAIYIER(H)
FcRL1 (B cell activation coreceptor)	(V)YSLVYHTPQVLEPAA(A)
Heat shock 70 kDa protein 13	(M)AEEYLGMPVANAVIS(V)
GAPDH (glycolysis)	(L)TGMAFRVPTPNVSVVD(L)
Myotubularin-rel protein 4 (phospho PIP3)	(L)DARSYTAAVANRAKGGG(C)
Aspartate aminotransferase	(T)KNLDYVATSIHEAVTK(I)
LPS-induced TNFa factor homolog	(M)NPPSYYTQPVPVPNAN(A)
Ly-9 (T cell adhesion)	(T)HLNTYDGSHTLRVSQS(V)
Semaphorin-4D	(K)GRKFVYAGSNSGVVQ(A)
Semaphorin-4D	(T)KWVRYNGPVPTPRPG(A)
Semaphorin-4D	(K)GRKFVYAGSNSGVVQAP(L)
Semaphorin-4D	(K)GRKFVYAGSNSGVVQAPL(A)
UDP-glucuronic acid decarboxylase	(E)VDQIYHLASPASPPN(Y)
E3 ubiquitin-protein ligase RNF13	(N)AQRAGYKAAIVHNVDS(D)
E3 ubiquitin-protein ligase RNF13	(N)AQRAGYKAAIVHNVDSD(D)
E3 ubiquitin-protein ligase RNF13	(Q)RAGYKAAIVHNVDS(D)
E3 ubiquitin-protein ligase RNF13	(Q)RAGYKAAIVHNVDSD(D)
MLV-related proviral Env polyprotein	(T)DAGKKASWDGPKVWGLR(L)
MLV-related proviral Env polyprotein	(A)GKKASWDGPKVWGLR(L)
Immunoglobulin superfamily member 7	(V)RPGGGPVSVELVGPRSHRLR(L)
Neutral alpha-glucosidase AB	(R)VPDVLVADPPTARLSVSGR(D)
Lysosomal alpha-glucosidase	(L)DVYVFLGPEPKSVVQ(Q)
K-D alpha chain	(E)GPEYWEEQTQRAKSDEQ(W)
L-D alpha chain	(A)GPHSMRYFETAVSRPGLG(E)

The list of peptides identified by MS/MS and the protein from which they belong. They are aligned by their I-A^b binding motif, with the major I-Ab binding motif in red, and all come from proteins that potentially could be processed and presented by MHCII. This list was generated from multiple MS/MS runs on at 4 biological samples.

Table 4: Comparing the Peptides Found in WT with the Mutant Samples

	WT peptides extracted	percent found in mutant	percent intensity
A64Q	219	79.9	96.5
R70A	219	84.5	97.1
T77A	219	84	95.3

A summary of the peptides found reproducibly in WT samples. Peptides were eluted from WT I-A^b and analyzed by mass spectrometry. Peptides with identical HPLC retention times that were present in three separate WT samples were identified. Data indicate the percentage of peptides that were also identified in duplicate runs isolated from βT77A, βR70A and αA64Q. Data are representative of multiple MS and MS-MS runs.

appearing only in the mutants, first a list of peptides that were found in duplicate mutant MS runs of the same sample was created. Any peptide in this list that also appeared in any of the three WT samples was also called present. Once again most of the peptides and nearly all of the intensity from the mutant samples were found in the WT runs (Table 5). If less stringent criteria were used, e.g. requiring presence in only two of the three WT or mutant I-A^b samples, even more shared peptides were identified among the samples. As this analysis does not identify which peptides are belonging to nested sets with a shared binding motif, I may be underestimating the similarity by considering differentially trimmed peptides as different peptides. Altogether, these experiments support the notion that the mutations of βT77A, βR70A and αA64Q MHC residues do not notably alter the repertoire of self-peptides bound to I-A^b. Therefore, I proceeded to introduce these mutations into the genome of C57BL/6 mice to test the effects of mutations in MHC on thymic selection *in vivo*.

Discussion

In this chapter I selected the mutations that I will extensively interrogate in thymic selection. I started by examining the solvent exposed residues on the α helices of the I-A^b that often interact with TCRs in crystal structures of TCR-pMHC complexes. Through sequence alignment of polymorphism and conservation, I chose 10 amino acids to test for mutation causing a reduction in mature T cell activation. From this mutational analysis, I chose A64Q on the α chain T77A and R70A on the β chain. α64A is invariant in mouse and human MHCII molecules and creates a docking “cup” for TCR Vβs that contain a tyrosine (Y) at position 48 of

Table 5: Comparing the Peptides Found in the Mutants with WT Samples

	mutant peptides extracted	percent found in WT	percent intensity
A64Q	152	98.7	99.2
R70A	228	87.3	91
T77A	180	91.7	97.5

A summary of the peptides found reproducibly in each of the mutant samples. Peptides were eluted from mutant I-A^b and analyzed by mass spectrometry. Peptides with identical HPLC retention times that were present in duplicate runs of the same sample were identified. Data indicate the percentage of peptides that were also identified in any of the WT runs. Data are representative of multiple MS and MS-MS runs.

CDR2 (43). βT77 is invariant in common mouse I-A and I-E alleles and in human HLA-DR and HLA-DQ alleles, and in TCR/MHC structures, it often is contacted by the TRAV CDR1 loop (43). βR70 is nearly monomorphic in all mouse I-A alleles (Table 1), but not conserved in mouse I-E alleles or in the MHCII alleles of other species. In nearly all published TCR/I-A structures it lies in the central region of the TCR footprint interacting with the TCR CDR3s and therefore, might be expected to influence somatic CDR3 selection during thymic selection, but perhaps not so much germline Vα and Vβ usage, so this mutation might act as a control for my future experiments. To confirm that any effect these mutations have on developing or mature T cells is a direct result of TCRs recognizing the mutation, MS and MS/MS analyses were performed on the peptide repertoire of these mutants and found there is no gross shift in peptide binding to MHC, as predicted from the crystal structures.

I-A^b is an obvious MHCII candidate as it is the only MHCII molecule in C57BL/6 mice, thus when looking at $CD4^+$ T cells I know the only MHCII they could interact with is I-A^b. In theory, however, not only could I have studied this on other I-A haplotypes, but I could have designed this study centered around I-E molecules or even any of the class I MHC molecules, but it would have been more difficult to distinguish between the confounding effects of multiple types of those MHC present in mice. Furthermore, in this study I only mutated amino acids that were not alanines to alanines and alanines to glutamines, so in cases where there was not a huge effect of the mutation on T cell responses, it is possible if I tried the entire panel of 20 amino acids, or maybe just larger and bulkier phenylalanine or tryptophan, I would have seen greater effects. Also, I never combined any

mutations on the same or opposite chains, which could have definitely led to larger effects as I would in theory disrupt more of the germline encoded interactions at once.

Choosing M12.C3 as the APC in these studies had both its advantages as well as disadvantages. Chosen for its ability to process and present antigens which allowed me to easily test bulk populations, it seemed like a good choice. However, the mixed molecule expression previous discussed coupled with an overall weak stimulation when compared to an APC like Chb complicated the stimulations. It is possible, however, that the weak stimulations provided by M12.C3 actually helped me visualize greater effects on T cell responses as they were more sensitive to my mutations. As briefly discussed earlier about different mutations altering mAb binding, the levels of MHCII on the surface of these cells were not as rigorously characterized as they might have been. Perhaps sorting on levels of transduction markers, rather than cloning these by limiting dilution, would have made it easier to generate more precise APCs for these assays.

The antigens chosen in these experiments had a wide variety of characteristics. The B5 peptide has only one predicted register for the peptide to bind MHC, so I thought these experiments would be more straightforward. The BNP peptide contained at least 2 different registers, which could be distinguished by a shorter version of the peptide used to stimulate. The shorter peptide was used and in fact did stimulate a smaller percentage of the bulk fusion, but I ended up only using the longer one for the analysis as it stimulated a larger percentage of the hybridomas. The whole proteins possess many epitopes to be presented by I-A^b so

it was a good tool to really explore the overarching effects of the mutations. A curious result in these experiments involved the MOG bulk hybridoma, because its cognate antigen is an autoimmune peptide. This bulk hybridoma was the only one tested that responded at all above background to the R70A mutation. With those hybrids being against a self-antigen, it is possible that they recognize MHC in a non-conventional manner as some other autoimmune TCRs do (101). It could be interesting as a future study to clone the bulk hybridomas specific against the MOG peptide that either respond to the R70A mutation or not, and explore the structural ways in which autoimmune TCRs can escape selection.

The use of CD69 as a readout for overall activation works as I applied it, but there may have been a better way to measure activation in these assays. It is actually a common problem when interpreting T cell activation data of this type, as it is often an assumption that the response, however it is measured, of individual cells occurs only because that cell was directly stimulated through its TCR. The bystander activation observed here shows that this assumption is obviously not the case. When I saw that the bulk hybridoma has a differential sensitivity to bystander activation when compared to a cloned hybridoma, the complexity of the system led me to abandon my initial goal of being able to glean more information about the cells that failed to respond to certain stimulations. If I had rigorously and systematically figured out what was causing the bystander activation, then I could have attempted to block it and truly only measure TCR mediated stimulations. This strategy would have allowed me to interrogate the T cell responses in the way I initially intended and could have yielded valuable information about certain TRAV families being most

sensitive to which mutations. Ultimately, because my main goal was to study the effects of these mutations on thymic selection and not in my *in vitro* system, I moved to the real experiment discussed in the next chapter.

CHAPTER IV

I-A^b MUTANT MICE

Introduction

In this chapter, I will show the data from the generation of I-A^b mutant mice expressing the T77A, R70A, and A64Q mutations. I will discuss the strategy and struggles by which the mice were generated. These mice will be shown to be an appropriate model in which to study the effects of these mutations on thymic selection. Additionally, the initial phenotypic characterization of T cells in these animals will be shown. In this characterization, the data highlighting the first effect of the mutations on T cell development will be presented.

Generation of Point Mutant Mice

For the aforementioned reasons, I chose T77A and R70A on the MHCβ side and A64Q on the MHCα to test in thymic selection. One roadblock I faced was how to put the mutations back in the thymus of a mouse. Because of the role of non-hematopoietically derived cells in expressing MHCII during thymic selection, these mutants cannot be introduced into the thymus as retrogenics, such as those expressing TCRs. Many transgenic animals have been constructed with an MHC expression vector (102) although the level of transgenic MHC expression is not always equivalent to the physiologic expression. Still valuable insight has been gained from some of these studies (62). In addition, other interesting MHC expression experiments have been conducted, such as expressing HLA-DM on the surface of cells and showing that an MHC structure mimic is not sufficient to positively select T cells (103). However, for my experiments subtle shifts in the

repertoire could yield very useful information about how the repertoire compensates for my mutations, and thus, getting the expression location and levels exactly physiological was crucial. To this end I changed the coding sequences of the genes in question, using zinc finger nuclease (ZFN) technology to generate knock-in point mutations directly in fertilized C57BL/6 eggs (104, 105). These experiments were done in collaboration with the National Jewish Health Mouse Genetics Core with Jennifer Matsuda and James Gross.

Custom ZFNs were designed to both *H2-Ab1* and *H2-Ab2* the genes that encode the α and β chains of the only MHCII molecule expressed in C57BL/6N mice (Sigma Aldrich, St. Louis, MO). To reduce off target effects, the ZFNs were designed to ensure that no other region of the mouse genome had fewer than five DNA base mismatches to the sequence targeted by the ZFNs. Custom ZFNs were first tested for activity in a Cel-I mismatch assay (Transgenomic, Omaha, NE), which measures double stranded breaks (DSB) repaired by error-prone non-homologous end joining (NHEJ) which changes the genomic DNA by making insertions and deletions at the site of the double stranded break. I first tested the custom ZFNs for activity by transfecting the plasmid encoding the ZFNs into JM8, an embryonic stem cell line, as well as mouse embryonic fibroblasts (MEFs). While optimizing the transfection could potentially have yielded a more strong effect, I was still able to see slight activity at the site of interest (Figure 5).

Now that I knew the ZFNs were active, I decided to try to introduce my mutations into mice. To do this, I would have to create a template for homology directed repair (HDR). This template would require three components to be most

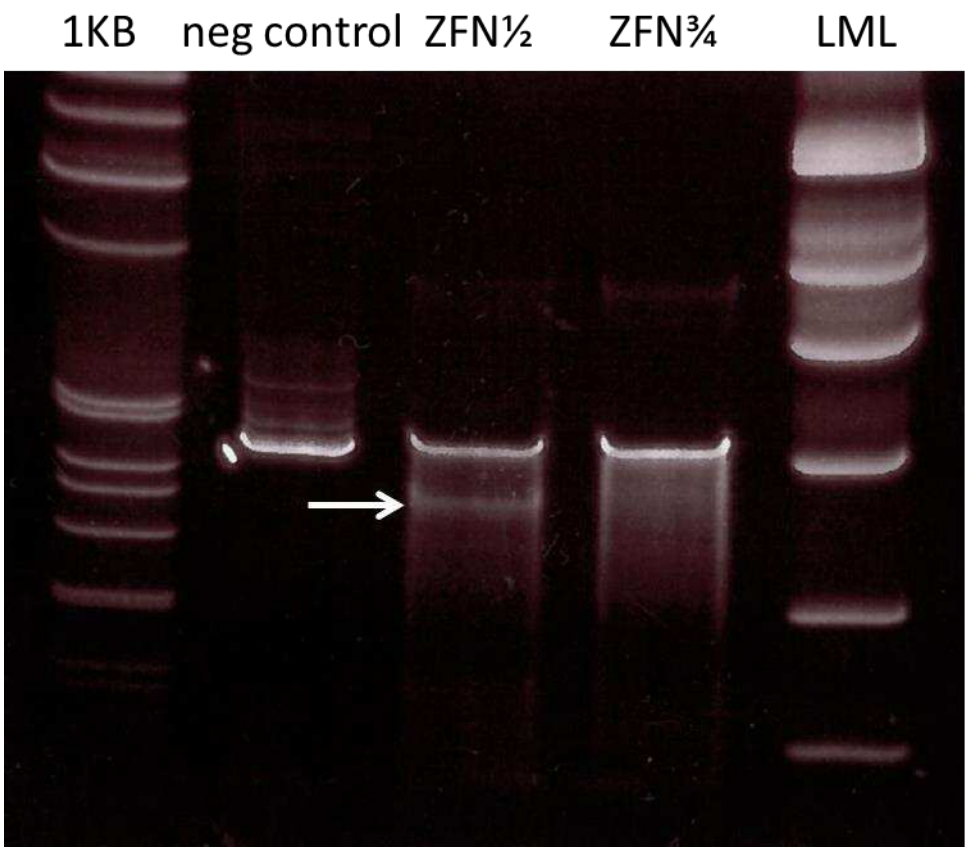

Figure 5: Cel-I Assay Shows ZFN Activity
Two of the ZFN pairs were transfected into MEFs and their activity was measured by mismatch detection using the Cel-I assay. Results are a measure of incorrect repair of the ZFN cut site by NHEJ events. The faint band showing activity is marked with an arrow for the first pair, ZFN1/2, while no activity was detectable in the second pair, ZFN3/4, above the negative (neg) control. This experiment is representative of two biological replicates and served only to provide confidence to test the ZFNs in the fertilized eggs, a more relevant setting for the intended construction of the mice. 1KB and Low Mass Ladder (LML) are DNA ladders confirming the correct size of the PCR fragment.

effective: the mutation of interest, a silent mutation to create a new restriction enzyme site for screening of pups, and a silent mutation that would disrupt the ZFN binding such that a subsequent insertion or deletion event caused by NHEJ would not occur after my mutation of interest had been introduced (Figure 6). Originally I attempted to use a single stranded DNA oligonucleotide simultaneously as other groups (106). Their results, however, were much more efficient than mine. I saw that my ZFNs were active, measured by deletion events around the ZFN cut site via sequencing. To confirm the screening process, I stained peripheral B cells from mice I identified as either knockouts, heterozygous, or WT, and their levels of MHC were as predicted (Figure 7). However, I only had one case of HDR, and that was only the ZFN-disrupting mutation and not the mutation of interest. After many failed attempts as well as declining birthrates, I decided to switch to a better established technique (107) for HDR. This technique involves the use of a double stranded DNA fragment with roughly 1000 bp of homology around either end of the target of the ZFN. Using this technique, all 3 point mutations of interest were generated. These data are summarized in Table 6. The method was surprisingly robust, with NHEJ events identified in nearly all of the mice and at least one chromosome with the correct mutation found in >10% of the mice overall.

Phenotypic Analyses of Mutant Mice

Now that the mice expressing these mutations had been generated, I could begin to assess the T cells in these animals. Mutant mice were crossed to wild type mice and then intercrossed to create mice homozygous for each of the three mutations. The first confirmation was that the levels of MHC are equivalent in these mutant

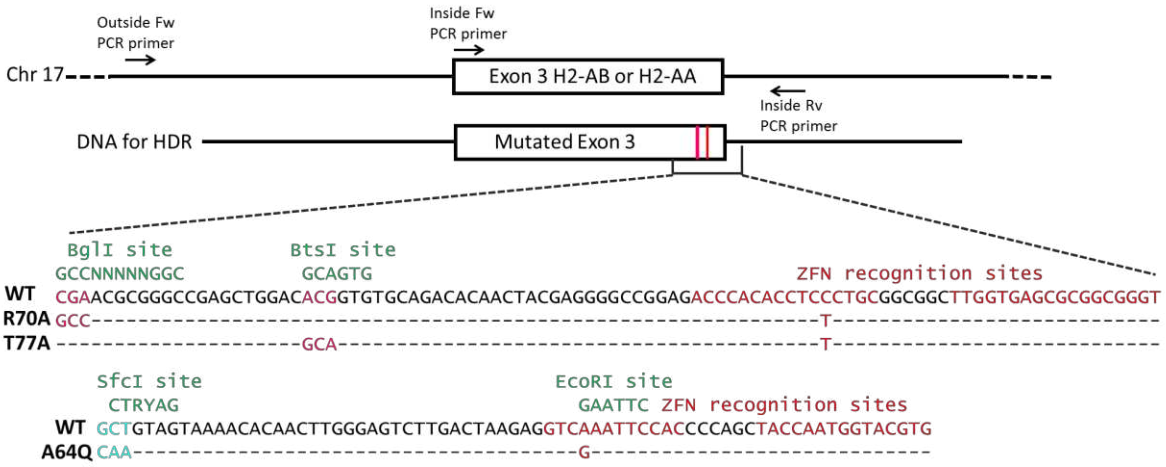

Figure 6: Schematic of HDR to Make I-A^b Point Mutants

A schematic for the strategy employed to generate point mutants is depicted. The point mutations were introduced in both the α and β chains of I-A^b on chromosome 17 of the mouse. In addition, two silent mutations were introduced: one to prevent further ZFN activity after introduction and another to introduce a novel restriction site for screening purposes.

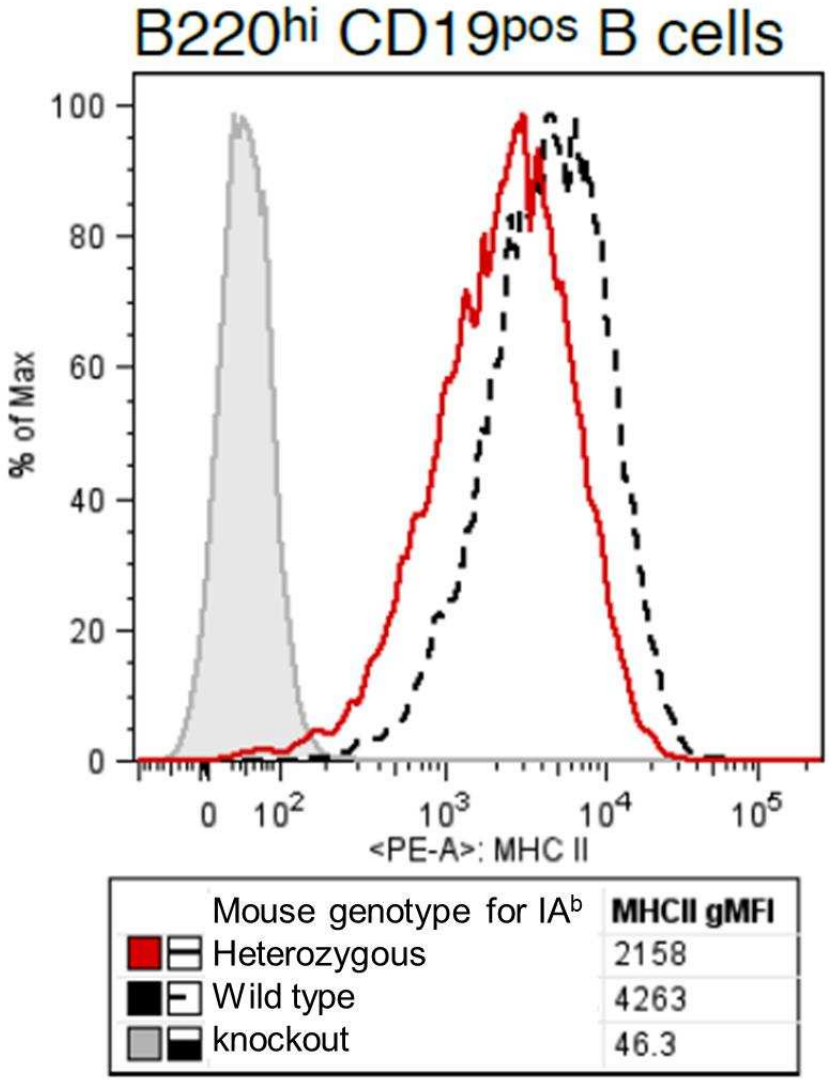

Figure 7: MHCII Expression Confirms Mouse Screening
Peripheral B cells ($B220^{hi}$ $CD19^{+}$) from mice that were WT, heterozygous, or deficient for $I\text{-}A^{b}$ were stained for surface MHC-II. Mice were identified as belonging to a given genotype by screening their genomic DNA. The gMFI of MHC-II on B cells from each genotype is also provided. This experiment was not repeated.

Table 6: Summary of Injections with Large DNA Fragment HDR

Table 6 Zinc finger targeting of fertilized eggs to produce mutated IAb genes				
Expected Mutation	Length of homology fragment used to target	Number of pups screened	Number of mutant IAb genes	Number of IAb genes with the desired mutation only
IAb alpha chain A64Q	2052	34	21	1
IAb beta chain R70A	1862	48	56	10
IAb beta chain T77A	1862	10	10	3
Note, in some cases breeding showed that mice that had more than one mutated gene were mosaics, that is one of the mutational events occurred at the embryonic 2 cell stage.				

A summary of the injections that were done to make all three mutant mice of interest is shown. The percentages highlight extremely efficient ZFN activity as well as high levels of HDR with the cut-plasmid template.

mutant mice as measured by the mAb M5/114 on peripheral cells, an I-A^b β chain specific antibody whose binding was not affected by any of the mutations when tested M12.C3 transductants but stained the mixed I-$A^{b/d}$ mixed molecule too intensely to be useful in those studies. The staining data in the mutant mice (Figure 8) show that the mutations had no effect on the expression of I-A^b on splenic B cells. In addition to confirming the expression of the mutant MHCII, I sought a functional readout that the mutations were actually expressed in these animals. To do this, I once again stimulated the B3K506 hybridoma (as shown in Figure 3) but instead of using M12.C3 cells expressing the mutations, I used splenocytes from the MHCβ mutant mice (Figure 9). As I expected, T77A mice showed a shift in the titration curve and R70A completely ablated the response, giving me more confidence that the mice were indeed expressing the mutations of interest.

Initially, I was expecting to see a reduction in CD4 single positive (SP) thymocytes in the T77A and A64Q mutant mice. At best, I hoped for a reduction in similar to what was seen in the TCRβ mutants (66). To determine whether any of the MHC mutations might have affected the development of $CD4^+$ T cells, the thymus of each mouse strain was analyzed by flow cytometry. No significant difference in the number of thymocytes in the double negative (DN), double positive (DP) and single positive (SP) populations was detected between the thymi of the mutant mice compared to WT mice (Figure 10). I used co-expression of CD5 and CD69 as markers of DP thymocyte activation during positive selection. Compared to WT mice, the size of this population was not changed in the βR70A and βT77A mutant mice but was significantly reduced in the αA64Q mutant mice (Figure 11).

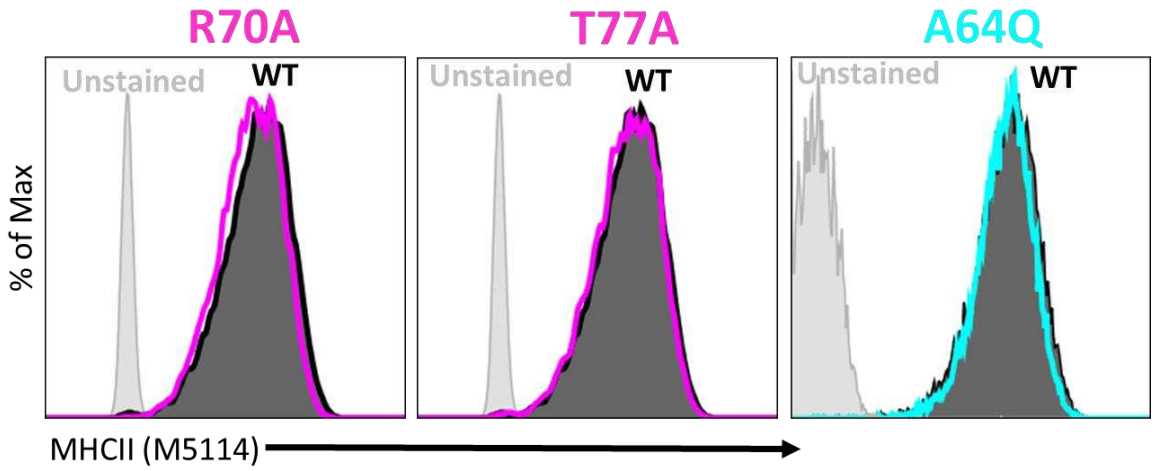

Figure 8: MHCII Levels in Mutant Mice Are Unchanged
MHC-II levels on peripheral non-T cells from both WT and mutant I-A^b mice are portrayed in a histogram. I-A^{b+} cells in the thymus also display a similar profile to the one depicted here. This experiment is representative of at least 7-10 mice stained in three biological replicate experiments.

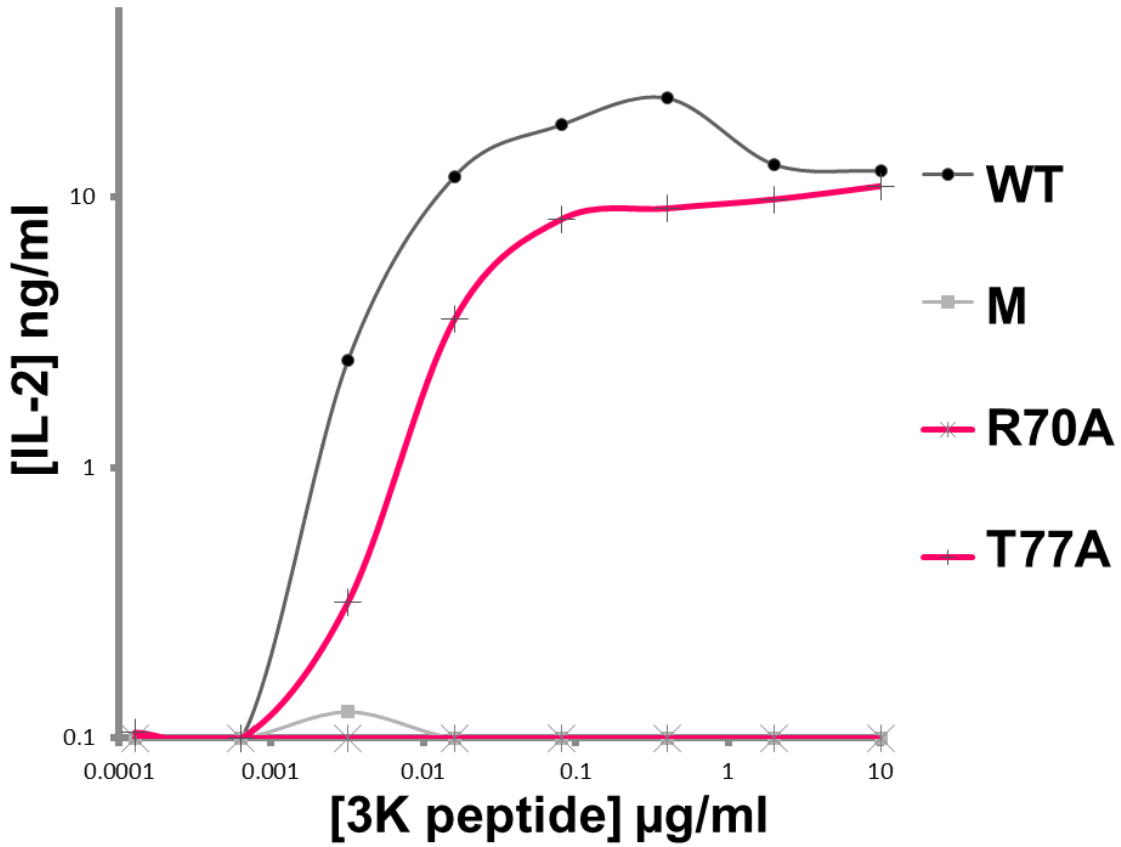

Figure 9: Splenocytes from MHCβ Mutant Mice Stimulate B3K506 with the Same Pattern as the Transduced APCs

B3K506 hybridomas were cultured with different amounts of 3K peptide along with splenocytes from the different WT or mutant MHC mice to generate a dose-response curve. Splenocytes from a H-2^f haplotype mouse (B10.M) were used as a negative control. IL-2 production by the hybridomas was used as a readout of activation. This experiment is n=1 and was not repeated, but provided confidence by recapitulating the pattern from an earlier experiment with the APC transductants (Figure 3).

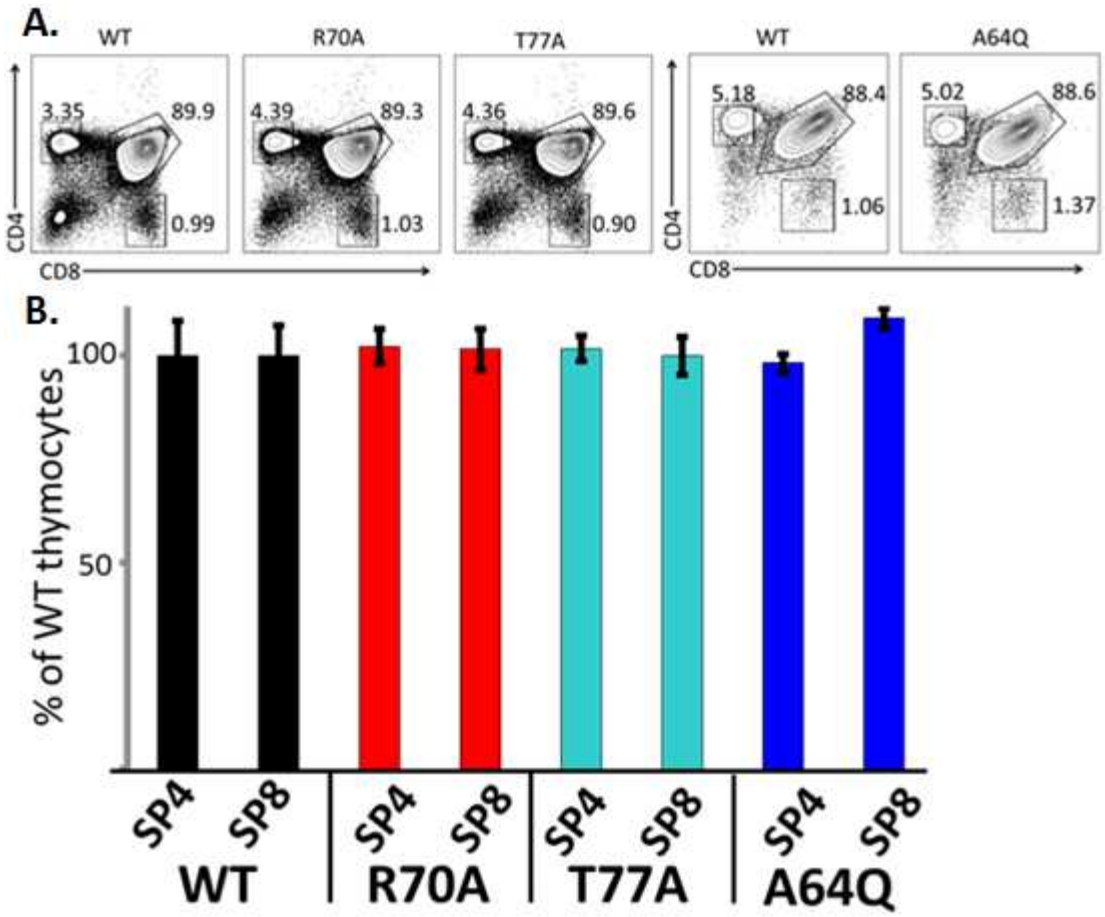

Figure 10: Thymic Bird of Mutant Mice is Similar to WT
Representative CD4/CD8 bivariate plots are depicted for thymocytes from the WT and mutant mice (top). The proportion of single positive CD4 and CD8 cells is quantified for each genotype and normalized to that found in WT mice (bottom). Data are aggregated from at least three independent experiments and 7-10 mice per group. A64Q experiments were conducted as separate experiments on different equipment but always normalized to WT age and sex matched controls.

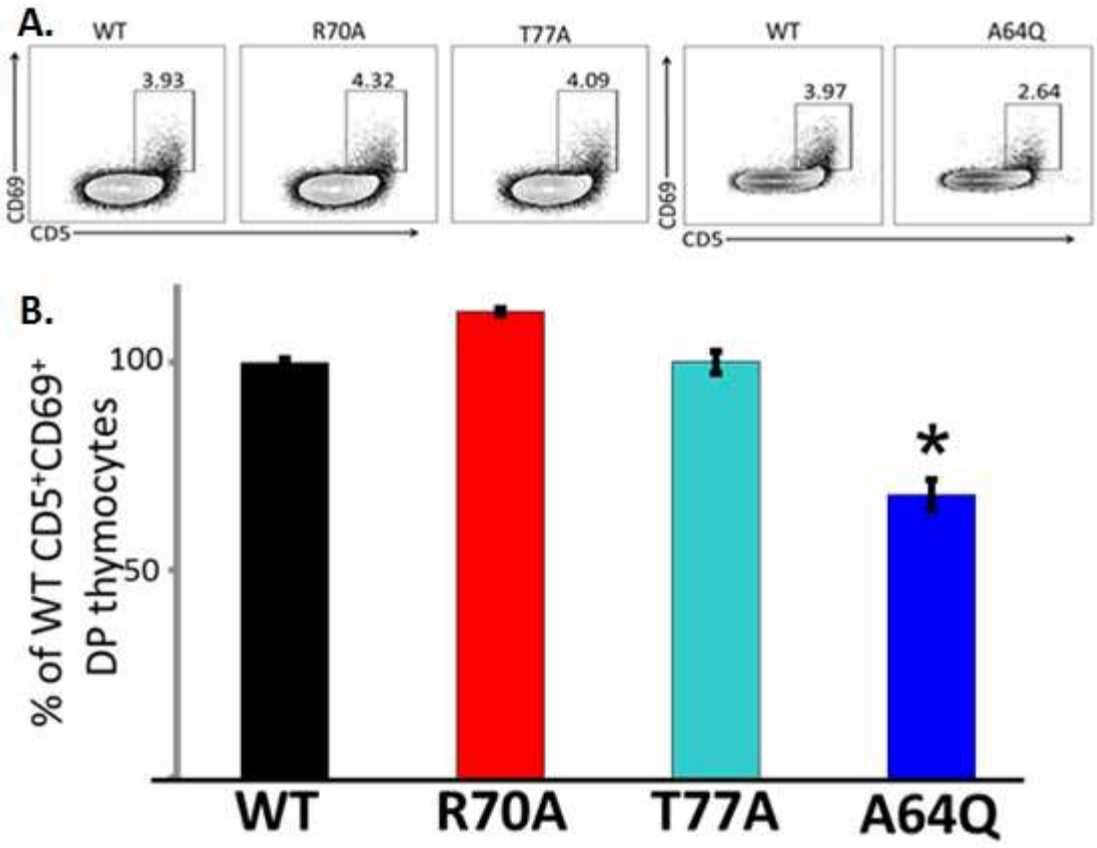

Figure 11: Activation of Thymocytes is Reduced in A64Q Mutant Mice
Representative staining of activated double positive thymocytes as measured by CD5 and CD69 coexpression is depicted for each genotype (top). The proportion of cells in each genotype which are $CD5^{+}CD69^{+}$ is quantitated and normalized to that found in WT mice (bottom). Data are aggregated from at least 7 mice per group over 3 separate experiments. The asterisk represents a p value < .05 for a one sample t test with true value 100.

This reduced activation in the thymus was an indication that perhaps the other mutations were also impairing selection and encouraged me to explore the mutant TCR repertoires more deeply.

Discussion

In this chapter the knock-in mutant mice that are the crux of this thesis were generated. To do this, I received ZFNs specific for the gene encoding I-A^b β exon 3 first and then later I-Ab α exon 3 on chromosome 17. I showed that when Itransfected the first pairs of ZFNs into MEF cells, this ZFN was active and causing NHEJ events around the cut site. I then devised a scheme to introduce the mutations of interest into single-cell embryos. This scheme involved a fragment of DNA, either and oligonucleotide at first, or later a fragment of large plasmid, to introduce the mutation through HDR. While the oligonucleotide was unsuccessful for directing the repair, I could measure ZFN activity in the embryo by NHEJ events causing mice born to either be heterozygous or complete knockouts for the MHCII. Upon switching to repair using a large DNA fragment as a template, all three mutants of interest were generated. These mutant mice expressed normal levels of MHCII upon homozygous breeding and stimulated a known hybridoma in a similar pattern as the transduced M12.C3 APCs. Thus, these mice are an appropriate system to study the effects of the mutations on the selection of CD4 T cells in the thymus.

Upon first examination, disappointingly, there is not a difference in the number or percent of single positive CD4 cells in the thymus, and, when measuring activation by CD5 and CD69, I can detect less of positively selecting signal being

recognized by single positive CD4 cells only in the A64Q mutant mice. However, this reduction was the first sign of the mutations affecting positive selection, and encouraged me to more deeply examine the TCR repertoire in all of the mutant mice.

Since generating these mice, other methods have appeared for nuclease directed introduction of genomic mutations. Examples are transcription activator-like effector nucleases (TALEN) (108), which function in a very similar manner as ZFNs, as well as the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein 9 (Cas9) system which results also in a DNA targeted break but is mediated by an RNA-guided Cas9 endonuclease (109). Both of these systems potentially could have offered me less expensive and time consuming ways to generate the knock-in mice. The CRISPR/Cas9 system has now been appropriated as the major genome modifying tool and is incredibly efficient (110).

Although I present the construction of all three mice simultaneously, actually the T77A and R70A mice were generated first, and then followed by the A64Q mutation some time later. This had to do with the order in which I received the ZFNs. I started with the MHCβ because I knew with only one set of ZFNs both lines of knock-in mice could be generated. While it took much troubleshooting to generate the T77A mouse, I knew that since I was using the same ZFNs and extremely similar method for HDR, I would be able to generate the R70A mouse shortly afterwards. Once I abandoned the oligonucleotide HDR method, the anticipated generation of the R70A mutation quickly followed. The A64Q mouse, on

the other hand, required an entirely new set of ZFN, as the I-A^b α chain is located about 60,000 base pairs away from the I-A^b β chain. Because of this proximity, it would actually be a challenge to create a mutation on both chains at the same time, as creating two DSB would likely cause a large deletion on the chromosome (111). Furthermore, interbreeding of mutant mice to generate a combined mutant is nearly impossible with this proximity. The best way to generate a combination mutant would be to inject the second mutation sequentially in fertilized eggs of the first mutation, something that might be worthwhile doing in the future.

Based on my initial mutational data combined with previous thymic studies (66), it was surprising when a greater effect on SP CD4 thymocytes in these animals was not observed upon my first phenotypic analyses. The data for the A64Q thymic activation was the first sign that in fact the thymocytes were receiving less of a positive selection signal, and this result indicated that the thymocytes developing in the mutants were seeing the mutation and somehow compensating for them. One of the questions that always arises regarding the evolutionary conservation of the TRBV13 family and Y48 is about its generality to other individual TRAV and TRBV families. This gave me the idea that since I was only making single point mutations to the surface of the MHC, perhaps there is a differential requirement of individual TRAV or TRBV subfamilies for the specific mutations. If only some of the TRAV or TRBV subfamilies are most sensitive to the mutations, then other families will still develop in the thymus and fill the SP CD4 niche. This hypothesis is what I explore in the next chapter.

CHAPTER V

ALTERED TCR REPERTOIRE IN MUTANT MICE

Introduction

In this chapter, the data from which I draw the strongest conclusions on how the T cells developing in the mutant mice are affected by the mutations are presented. To accomplish this analysis, next-generation sequencing techniques will be applied to both TCRα and TCRβ chains. Extensive data sets will be deeply analyzed to justify the conclusions drawn. Finally, the functional relevance of the differences observed will be addressed as well an interpretation of what these findings mean for the evolutionary conserved interaction of TCRs and MHCs.

Clues to an Altered TCRα Repertoire

If there is very little change in T cell numbers, how are the mutations affecting the TCR repertoire? One answer might be that the TCR repertoire has shifted in the T77A and A64Q mice. To address this questions, I first started to look more deeply at the MHCβ mutants, R70A and T77A, and their effect on the TCRα repertoire. Since the germline portions of the TCRs interacting with I-A^b βT77 were predicted to be those of TRAV CDR1s, I predicted that the TRAVs used in the mutant mice would be more affected by these mutations than the TRBVs. Therefore, I compared TRAV usage in the βT77A mice to that in the WT mice, using the βR70A mutant mice as a possible control because of its prominent interaction with randomly generated CDR3 regions. Anti-TRAV staining with the four available anti-TRAV mAbs (TRAV14, TRAV9, TRAV12, TRAV4) revealed a significant reduction in TRAV14 usage in the mutant mice (Figure 12). As expected, this reduction was seen only in CD4 T cells,

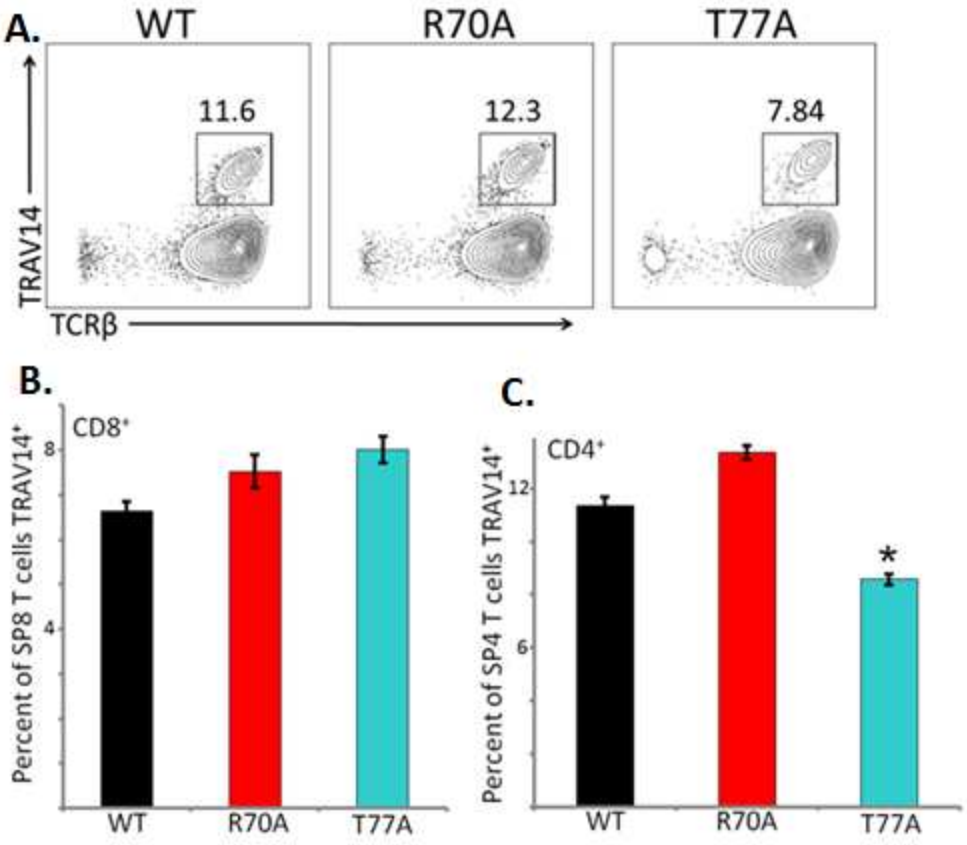

Figure 12: TRAV14 Shift in T77A Mice

A) Representative analysis via flow cytometry of TRAV14 expression in spleen $CD4^+$ T-cells from WT, βR70A, or βT77Q mice. B) Frequency of $TRAV14^+$ cells in $CD8^+$ or $CD4^+$ (C) T-cells in the spleen of the indicated mice. Data are representative of 3-4 independent experiments containing 7-10 mice per group. Error bars indicate S.E.M. and the asterisk represents a p value < .05.

not CD8 T cells. This TRAV14 shift was the first indication of an altered TCR repertoire in the T77A mice. However, this analysis was limited by the small number of anti-TRAV-specific mAbs available. Moreover, the TRAV antibodies that are available might not distinguish between the subfamily members of each TRAV family. For example, the TRAV14 family includes 9 subfamily members, some of which differ from each other, crucially for my studies, in their CDR1 and/or CDR2 amino acid sequences.

TCRα Sequencing in MHCβ Mutant Mice

Shifts in Family and Subfamily Usage in the T77A Mutant

To overcome this reagent limitation, I examined the TRAV repertoires of the MHCβ mutant mice in greater detail using deep sequencing. All deep sequencing was done in collaboration with the National Jewish Health Center for Genes, Environment and Health department. Naïve CD4 T cells were isolated from the spleens of the animals (Figure 13), in collaboration with the National Jewish Health Flow Cytometry Core, and RNA and then cDNA were produced from the cells. As described in the Material and Methods, I used a set of forward primers specific for the TRAV families and common Cα reverse primers to generate a diverse PCR product encoding the TRAVs present in the naïve CD4+ T cells from each strain of mice. Each of these fragments was sequenced with high-throughput methods. Using in-house developed software, short sequences were filtered out and the TRAV and TRAJ genes as well as the CDR3, germline and non-germline portion, used by each sequence were determined.

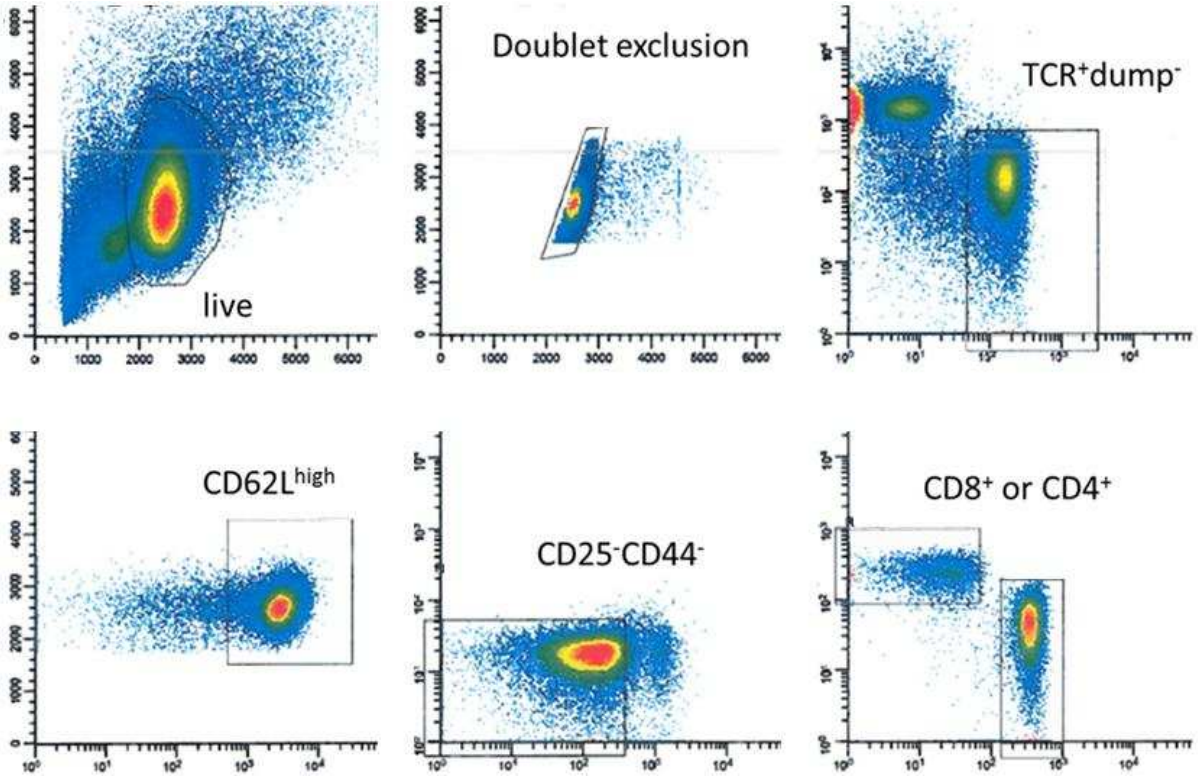

Figure 13: Sorting Naïve T Cells
A representative gating strategy for sorting naïve T cells out of the spleen of the WT and mutant mice. Live single cells that express a TCR and high levels of CD62L but negative for markers not expressed on αβ T cells (B220, F4/80, CD11b, γδTCR) and markers of antigen experience (CD25, CD44) were sorted based off of CD4 or CD8 into two separate populations from which I extracted the RNA to make cDNA.

To look for the effects of the mutations on the frequency with which different TRAVs are used, I compared TRAV family and TRAV subfamily usages among the mice. First, I looked at the average usage of the 20 TRAV families present in the mice (Figure 14) in the WT vs. mutant mice. Using the DESeq2 package (Love, Huber et al. 2014) that is often implemented in comparing mRNA expression in different cell populations, I discovered that there were not significant differences in frequency of use between the WT and R70A mice. However, there were reproducible and statistically significant differences in TRAV usage between the T77A and WT mice (Table 7). All statistical application of DESeq2 was done in collaboration with James Crooks in the National Jewish Health Biostatistics Department. The TRAV 3, 6, and 11 families were used more frequently and TRAV 5 and 7 families were used less frequently by T cells in T77A vs. WT mice. TRAV14 is trending towards being reduced, which would have confirmed the mAb staining results, but requires a deeper analysis to find statistical significance.

The TRAVs in C57BL/6 mice include a number of very closely related sequences; these have been designated subfamily members of larger families. Because of multiple duplications within the mouse TRAV locus, in C57BL/6 mice families can contain 1 to 15 related subfamily members of varying degrees of dissimilarity for a total of 109 TRAV elements. Each family is denoted by a number, e.g.TRAV1, TRAV6, etc. and a second number is used to denote family submembers, e.g. TRAV6-1, TRAV6-2, etc. However, genomic sequencing shows that portions of the TRAV locus in C57BL/6 mice have been duplicated twice. Because of this, subfamily members in one duplication are more like those in

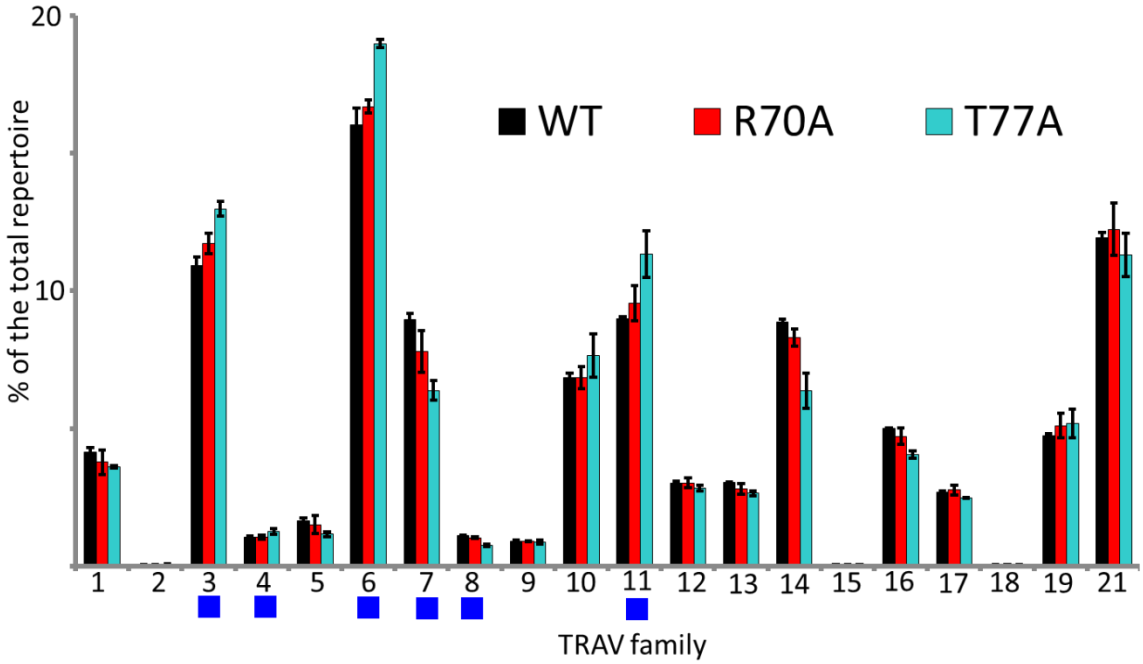

Figure 14: TRAV Family Usage in WT, R70A, and T77A TCR Repertoires
The overall TRAV family usage averaged from 3 individual runs of each genotype is represented in this bar graph. The error bars represent S.E.M. and the blue squares represent statistical significance (adjusted $p<.05$) as calculated by DESeq2 and shown in Table 7 for differences between WT and T77A. There are no statistical differences between WT and R70A on the family usage level.

Table 7: DESeq2 Statistics on T77A vs. WT Family Usage

	Log2 Fold Change	pvalue	Adjusted p value
TRAV01	0.089858	0.485233	0.671861
TRAV02	-0.1862	0.13826	NA
TRAV03	-0.29714	0.005784	0.026497
TRAV04	-0.29106	0.009547	0.03437
TRAV05	0.301103	0.037215	0.089454
TRAV06	-0.28288	0.01589	0.04767
TRAV07	0.332667	0.005888	0.026497
TRAV08	0.363811	0.004768	0.026497
TRAV09	-0.01693	0.881244	0.984997
TRAV10	-0.19876	0.105208	0.189374
TRAV11	-0.35039	0.003918	0.026497
TRAV12	-0.00261	0.977252	0.984997
TRAV13	0.094294	0.365829	0.548744
TRAV14	0.282762	0.039757	0.089454
TRAV15	-0.00233	0.984997	0.984997
TRAV16	0.184798	0.080675	0.16135
TRAV17	0.012109	0.916963	0.984997
TRAV18	-0.10668	0.486272	NA
TRAV19	-0.1593	0.282551	0.462355
TRAV21	-0.01523	0.915603	0.984997

The statistical output of DESeq2 on the family usage between T77A and WT repertoires as described in the Methods and Materials.

another than they are like other subfamily members. Thus, here, I will describe each TRAV by numbers denoting its family and subfamily, and also by a letter to describe whether its gene is located in what is thought to be the primordial set of TRAV genes (letter A) rather than the first (letter D) or second (letter N) duplication, for example TRAV6-5A, TRAV6-5D and TRAV6-5N (112). The sequences of these duplicated genes can sometimes be identical and therefore indistinguishable by RNA sequencing or sometimes quite divergent particularly in their CDR1 and CDR2 regions (43).

To find out if TRAV subfamily use differed between the different mice, I compared TRAV subfamily usage for all of the 88 TRAVs I could distinguish with the sequencing from the CD4+ T cells in the WT, βT77A, and βR70A mice. The data for all of the individual TRAV subfamily genes are contained in Figures 15 and 16 for each individual mouse as part of a heat map (green-high usage to yellow-medium, and red-low usage). The heat maps are ordered by the fold change (Figure 15-positive fold change, Figure 16-negative fold change) between T77A and WT such that the most reduced subfamilies in the mutant are located at the top and most increased are at the bottom with statistical significance calculated by DESeq2 denoted by blue squares. TRAVs underrepresented in the βT77A mice were 7-6A, 7-6D, 7-6N, 8-1AD, 8-2D, 12-1AN, 13-2AN and 14-3A. TRAVs over represented in the βT77A mice were 3-3A, 4-4D, 6-2A, 6-3ADN, 6-4A, 6-4D, 6-6N, 6-7A, 9-2D and 11-1AD. Differences between WT and T77A for all the subfamilies with their significance scores can be found in Table 8. Many of these subfamily differences account for the differences in overall family usage in Figure 14.

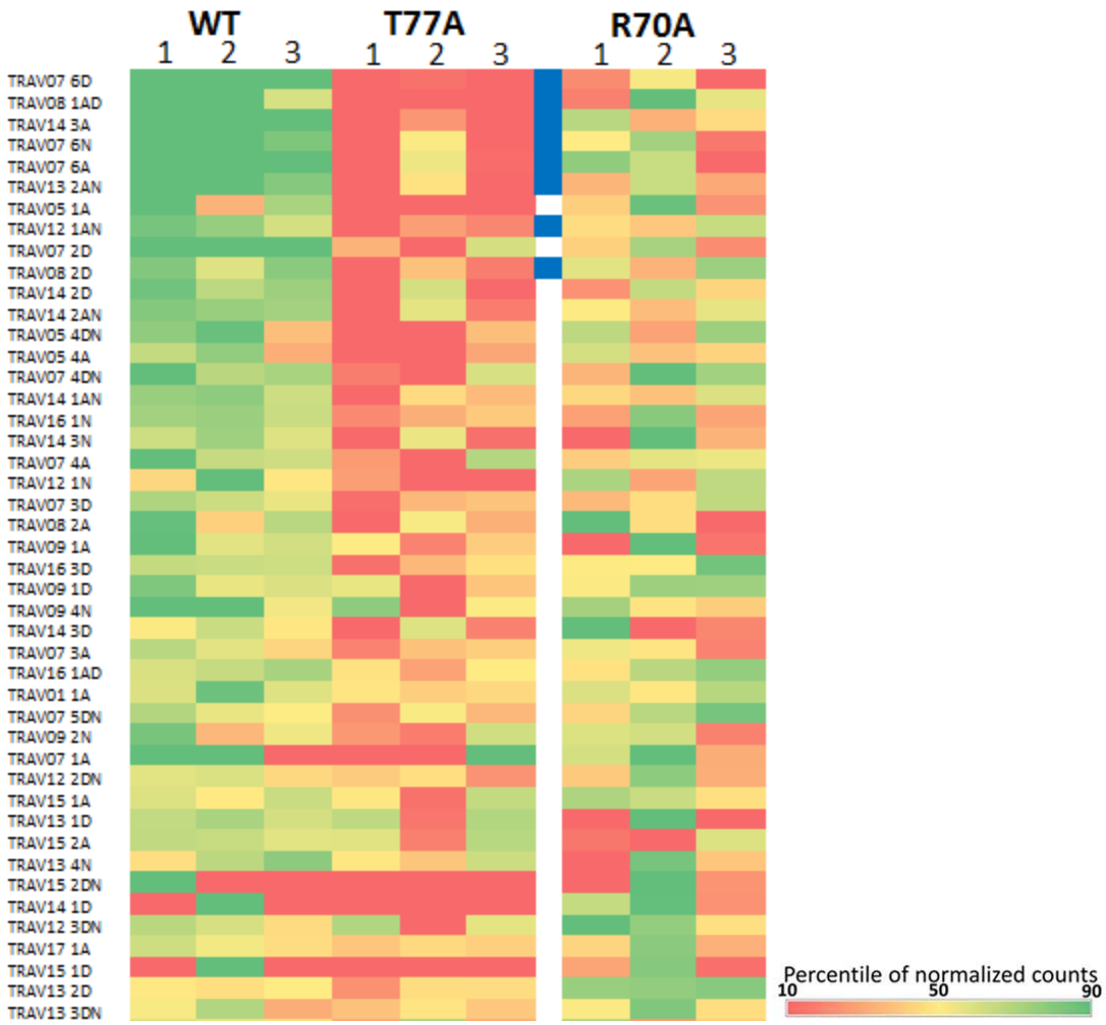

Figure 15: TRAV Subfamilies with a Positive Fold Change between T77A and WT

A heat map of subfamilies with a positive fold change between T77A and WT across all samples is shown. Green, yellow, and red indicate high, medium, and low usage, respectively. Statistical significance is indicated by a blue square for adjusted p <.05 between WT and T77A mutant mice. Statistical output in found in Table 8.

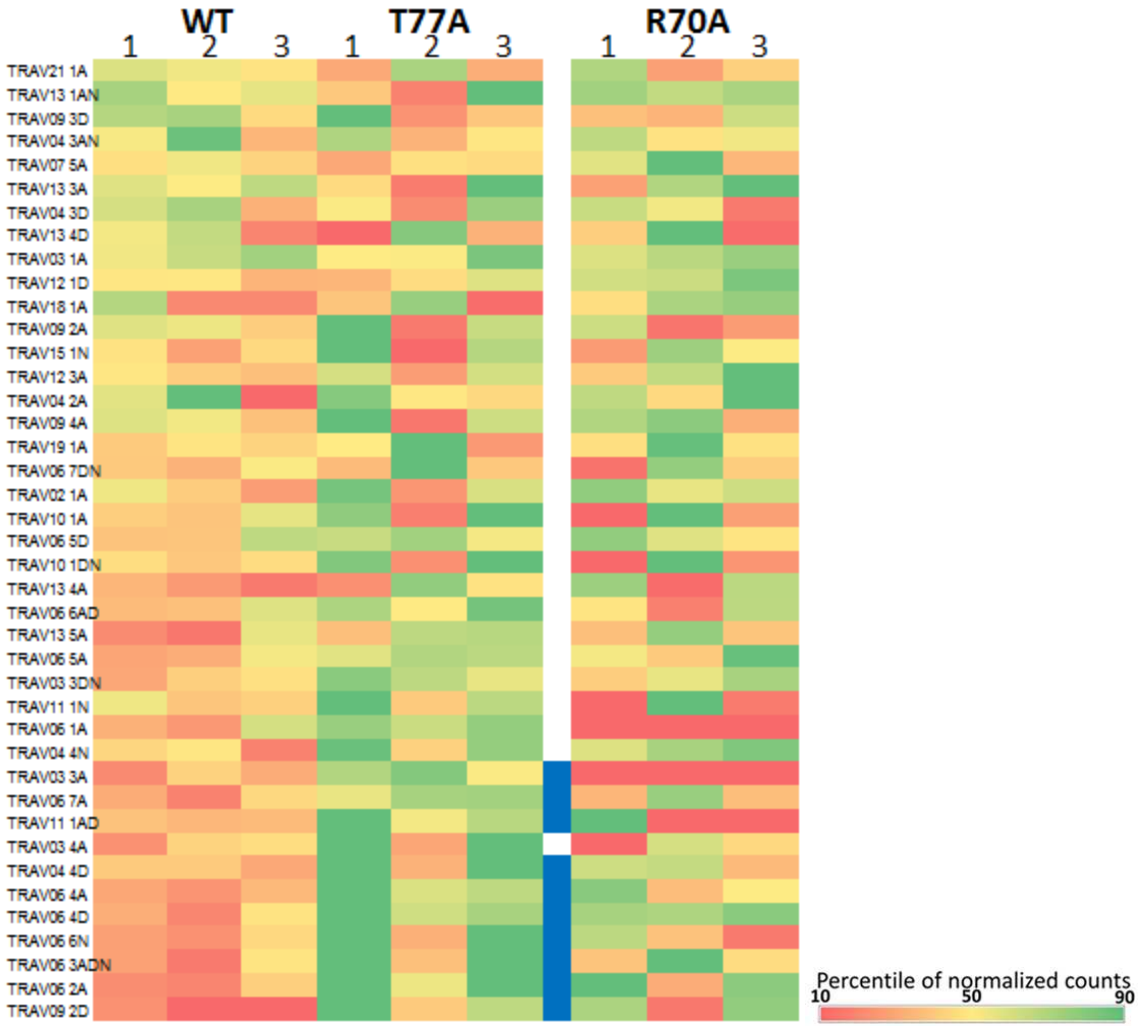

Figure 16: TRAV Subfamilies with a Negative Fold Change between T77A and WT

A heat map of subfamilies with a negative fold change between T77A and WT across all samples is shown. Green, yellow, and red indicate high, medium, and low usage, respectively. Statistical significance is indicated by a blue square for adjusted p <.05 between WT and T77A mutant mice. Statistical output in found in Table 8.

Table 8: DESeq2 Statistics on T77A vs. WT Subfamily Usage

Subfamily	Log2 Fold Change	Adjusted P value
TRAV07 6D	0.91	0.00020
TRAV08 1AD	0.65	0.00002
TRAV14 3A	0.61	0.00010
TRAV07 6N	0.51	0.01650
TRAV07 6A	0.50	0.04788
TRAV13 2AN	0.45	0.02961
TRAV05 1A	0.41	0.11099
TRAV12 1AN	0.40	0.01353
TRAV07 2D	0.37	0.10716
TRAV08 2D	0.36	0.04788
TRAV14 2D	0.36	0.11099
TRAV14 2AN	0.36	0.09718
TRAV05 4DN	0.34	0.11099
TRAV05 4A	0.33	0.11099
TRAV07 4DN	0.33	0.09718
TRAV14 1AN	0.26	0.12138
TRAV16 1N	0.25	0.07499
TRAV14 3N	0.25	0.19229
TRAV07 4A	0.24	0.27330
TRAV12 1N	0.24	0.38439
TRAV07 3D	0.21	0.18941
TRAV08 2A	0.21	0.33896
TRAV09 1A	0.19	0.28620
TRAV16 3D	0.19	0.25871
TRAV09 1D	0.17	0.34018
TRAV09 4N	0.15	NA
TRAV14 3D	0.13	0.51305
TRAV07 3A	0.12	0.46563
TRAV16 1AD	0.12	0.46563
TRAV01 1A	0.10	0.63847
TRAV07 5DN	0.10	0.63847
TRAV09 2N	0.09	0.76323
TRAV07 1A	0.09	NA
TRAV12 2DN	0.06	0.77747
TRAV15 1A	0.05	0.90749
TRAV13 1D	0.05	0.92264
TRAV15 2A	0.04	0.95188
TRAV13 4N	0.04	0.95188
TRAV15 2DN	0.02	NA
TRAV14 1D	0.02	NA
TRAV12 3DN	0.02	0.96543
TRAV17 1A	0.02	0.96543
TRAV15 1D	0.02	NA
TRAV13 2D	0.01	0.96543
TRAV13 3DN	0.00	0.99229

Table 8: DESeq2 Statistics on T77A vs. WT Subfamily Usage Cont'd

TRAV21 1A	-0.01	0.97915
TRAV13 1AN	-0.01	0.96543
TRAV09 3D	-0.01	0.96543
TRAV04 3AN	-0.01	0.96543
TRAV07 5A	-0.02	0.96543
TRAV13 3A	-0.02	0.96543
TRAV04 3D	-0.02	0.96543
TRAV13 4D	-0.04	0.96308
TRAV03 1A	-0.05	0.82313
TRAV12 1D	-0.09	0.61201
TRAV18 1A	-0.11	0.69596
TRAV09 2A	-0.12	0.61108
TRAV15 1N	-0.12	0.66099
TRAV12 3A	-0.13	0.46563
TRAV04 2A	-0.15	0.61201
TRAV09 4A	-0.16	0.48289
TRAV19 1A	-0.17	0.48289
TRAV06 7DN	-0.18	0.42806
TRAV02 1A	-0.19	0.34909
TRAV10 1A	-0.19	0.33728
TRAV06 5D	-0.19	0.28620
TRAV10 1DN	-0.23	0.19505
TRAV13 4A	-0.25	0.28620
TRAV06 6AD	-0.26	0.11099
TRAV13 5A	-0.27	0.19146
TRAV06 5A	-0.28	0.11554
TRAV03 3DN	-0.29	0.10319
TRAV11 1N	-0.30	0.13234
TRAV06 1A	-0.31	0.07275
TRAV04 4N	-0.32	0.07275
TRAV03 3A	-0.36	0.04078
TRAV06 7A	-0.36	0.02961
TRAV11 1AD	-0.36	0.04587
TRAV03 4A	-0.38	0.06653
TRAV04 4D	-0.39	0.02911
TRAV06 4A	-0.41	0.02911
TRAV06 4D	-0.43	0.01761
TRAV06 6N	-0.46	0.03018
TRAV06 3ADN	-0.50	0.01353
TRAV06 2A	-0.50	0.00117
TRAV09 2D	-0.54	0.00325

The statistical output of DESeq2 on the subfamily usage between T77A and WT repertoires as described in the Methods and Materials is shown in this table.

Richness and Diversity of the TCRα Repertoires

Sequencing identified not only the TRAV families and subfamilies used with the WT and mutant mice, but also the complete sequences of the TCRα domains, including the TRAVs, TRAJs and the somatically generated CDR3α regions. Thus I analyzed the diversity of the TCRα sequences among the naïve splenic CD4 T cells in the WT and mutant mice in several different ways. First, I examined the properties of the overall TRAV-CDR3-TRAJ repertoires.

Perhaps while the same number of T cells is being selected, there is a reduction in the diversity of the repertoire because fewer TCRα chains are used. One way I analyzed the repertoire was to explore what is known as species richness and diversity, or the number of TCR sequences and distribution in the sequencing runs. It is important to keep in mind that each sequencing run is from an individual mouse, and the number of total sequences collected is a result of the efficiency of the sequencing run and might not represent the TCRα repertoire extracted from the sorted cells. For this reason, when comparing populations of different sizes, I first referred to ecological measures that deal with sampling populations of species in the wild.

The first of these measures is the species accumulation curve, which visually compares the richness and diversity in a population, and was first used by Fisher et al. in 1943 (113). To do this I took a random sampling of the population along the X-axis and show on the Y-axis whether each sequence included adds a unique sequence to the total number of unique sequences (Figure 17). This curve should plateau as the data nears saturation of all sequences present in the cDNA sample.

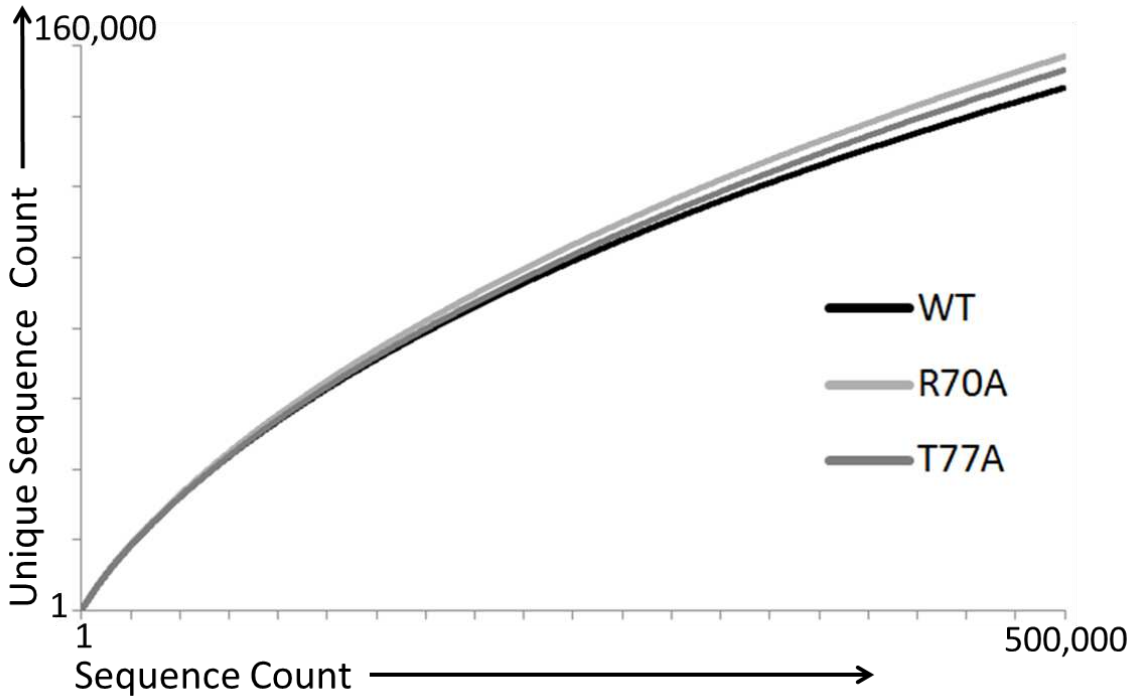

Figure 17: Species Accumulation Curves Reveal Similar Diversity and Richness of the TCRα repertoires

The average of the species accumulations curves for the different repertoires. The shape of the curves indicates similar species richness and diversity between the populations. The lack plateau in the slope suggests that the populations sampled in the sequencing runs are not saturated. Averages are representative of the individual curves and there are no differences between the samples.

Sequences of the TCRαs from the three types of mice have similar curves and do not plateau even after analyzing 500,000 randomized sequences. Therefore, the WT, βR70A and βT77A mice all express similarly large, diverse TCRα repertoires in their mature, naïve splenic T cells, which my sequencing does not completely saturate.

Comparing the WT and Mutant Repertoires

Next I wanted to ask how different each of the MHCβ mutant repertoires is from the WT. The key in these analyses was to ensure that since I am sampling a snapshot of an extremely large repertoire in all of these animals, I needed to make sure that analyses were reproducible. Also, because each sample produced a different number of sequences, simply looking at the percent overlap one way and then the other way is insufficient. There are many different measures of similarity or dissimilarity, and it was important to use ones that specifically answer the questions that I was trying to ask. Furthermore, I needed to address the similarity of abundance data, including multiple sequences of the same TCR, and count data, just counting unique TCRs. I began my analysis with some standard and easily calculable indices. To do this for count data, I used the Jaccard similarity index, named for Paul Jaccard, a Swiss botanist who coined the term *coefficient de communauté* in 1901 (114). The Jaccard similarity index is calculated by dividing the intersection of the sample sets over the union. For abundance data, this becomes more complicated as a TCR can be found many times in one sample but only rarely in another. Again I turned to ecologists who have been dealing with similar, albeit smaller, datasets for some time. Anne Chao et al. (115) use an

interesting and appropriate probabilistic approach to the Jaccard index. The Chao-Jaccard index produces a number between zero and one that is directly related to, but not exactly, the probability that if you chose a TCR in population A, that TCR will be found in population B.

When I analyzed the data on a whole of all the runs of TRAV sequences, I noticed that there was just as little similarity between repeated WT mice as there was for the mutant whether I compared the count data (Table 9) or abundance data (Table 10). This really told me something about the size of the population I was sampling and hinted that possibly deeper analyses could be required.

However, a different type of accumulation curve shows that this large repertoire is not randomly dispersed, i.e. the frequency of each sequence is not determined by a simple Poisson distribution (Figure 18). Despite the lack of saturation, the average frequency of any given unique sequence in the WT samples was about 5, but ranges from 1 to over 10,000. Using this frequency, I constructed a Poisson-predicted accumulation curve that predicts the proportion of total sequences that should accumulate as I added sequences that occur from 1 to 20 times. This curve predicts that, if TCRα usage is Poissonian, I should account for nearly all of the sequences by the time I include those that occur 15 times or less, but the experimental accumulation curve generated from the sequencing data shows that these sequences account for only ~50% of the total sequences. Likewise, there were more sequences found fewer than three times than predicted by the Poisson curve. Similar results were seen with the data from the mutant mice. Thus, despite the great diversity of sequences, their frequency was not as predicted by a Poisson

Table 9: Jaccard Similarity Index Comparing Unique Sequences

	WT1	WT2	WT3	R70A1	R70A2	R70A3	T77A1	T77A2
WT2	0.248							
WT3	0.233	0.264						
R70A1	0.250	0.252	0.229					
R70A2	0.237	0.236	0.222	0.239				
R70A3	0.231	0.231	0.221	0.245	0.221			
T77A1	0.233	0.235	0.219	0.246	0.233	0.221		
T77A2	0.236	0.239	0.230	0.251	0.226	0.239	0.231	
T77A3	0.242	0.243	0.228	0.252	0.243	0.230	0.247	0.239

The Jaccard similarity indices comparing every sample to every other sample are shown in this table. With all the numbers being in the same range, this indicates that on this level of analysis the populations on an overall TCRα level, the WT and mutants are indistinguishable from each other by this comparison.

Table 10: Chao-Jaccard Similarity Index Comparing Total Sequences

	WT1	WT2	WT3	R70A1	R70A2	R70A3	T77A1	T77A2
WT2	0.524							
WT3	0.498	0.535						
R70A1	0.561	0.561	0.528					
R70A2	0.501	0.499	0.470	0.536				
R70A3	0.511	0.510	0.486	0.565	0.487			
T77A1	0.518	0.520	0.491	0.565	0.510	0.512		
T77A2	0.515	0.515	0.491	0.570	0.489	0.527	0.517	
T77A3	0.521	0.519	0.492	0.564	0.514	0.512	0.540	0.517

The Chao-Jaccard similarity indices comparing every sample to every other sample are shown in this table. With all the numbers being in the same range, this indicates that on this level of analysis the populations on an overall TCRα level, the WT and mutants are indistinguishable from each other by this comparison.

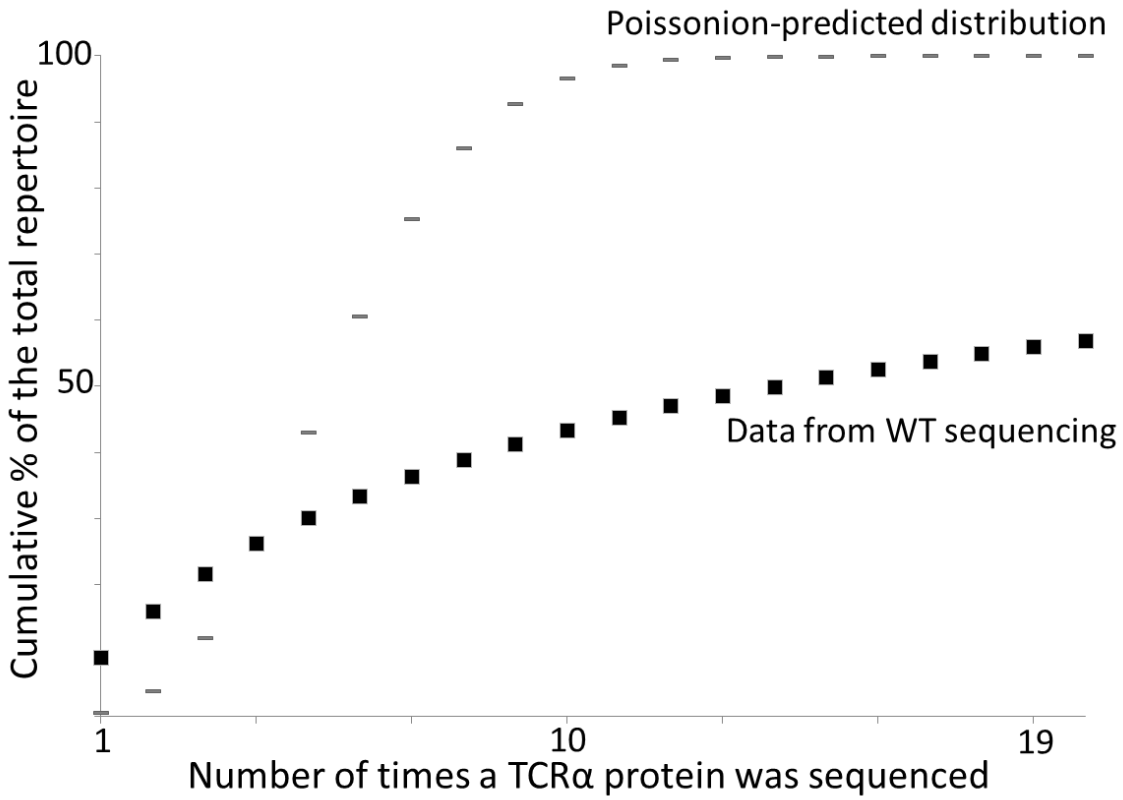

Figure 18: Poissonian-Predicted vs. Actual Data Distribution
A comparison of the of the Poissonian-predicted distribution and the actual sequencing data for one of the WT runs based off of the number of repeats in that run is shown on this graph. A Poisson distribution would predict nearly all of the total sequences found less than 15 times, but half of the actual data has not been accounted for at that point. All other samples differ from their Poissonian-predicted curves in a similar fashion. This is important as a Poissonian-predicted curve would indicate random sampling of a population.

distribution, a feature shared with previous repertoire analyses of different human T cell populations (116). While some of these results might be attributable to uneven efficiencies during the PCRs with the cDNA templates, it is likely that both thymic and peripheral selective pressures also contributed.

To get a better feel for the overlap between the populations, I created a ranking based off prevalence in the WT sample. In Figure 19, I aligned all the TCRs that were sequenced at least once in each of the WT runs, and plotted them in the rank order based on average percentage in each run. The error bars on the graph for WT TCRαs are S.E.M., and indicate a tight grouping of these reproducibly found α chains. I then plotted these TCR sequences for each of the MHCβ mutants in the same order as in the WT, to give a feel of where the overlapping chains lie, and possibly yield clues as to why the similarity indices did not show overall differences between the samples. The graphs are broken in to three sections to zoom in to see TCRs of different frequency. In the first graph, the top 1000 TCRαs are present in all genotypes, a feature that is weighted heavily in the Anne Chao-Jaccard similarity index. In the second and third graphs, TCRαs start to fall out of the mutant populations indicating that they might have been sensitive to the mutations made. The extreme variability in the third graph would be weighted heavily in the Jaccard similarity index, and probably accounts for the variable scores seen in Table 9. Overall, these graphs helped explain the similarity index results and led me to look for deeper statistical analyses of the overlap in these samples.

This deeper statistical method was to apply DESeq2 to two types of analyses on the total TCRα (combination of TRAV subfamily, TRAJ and CDR3) sequences.

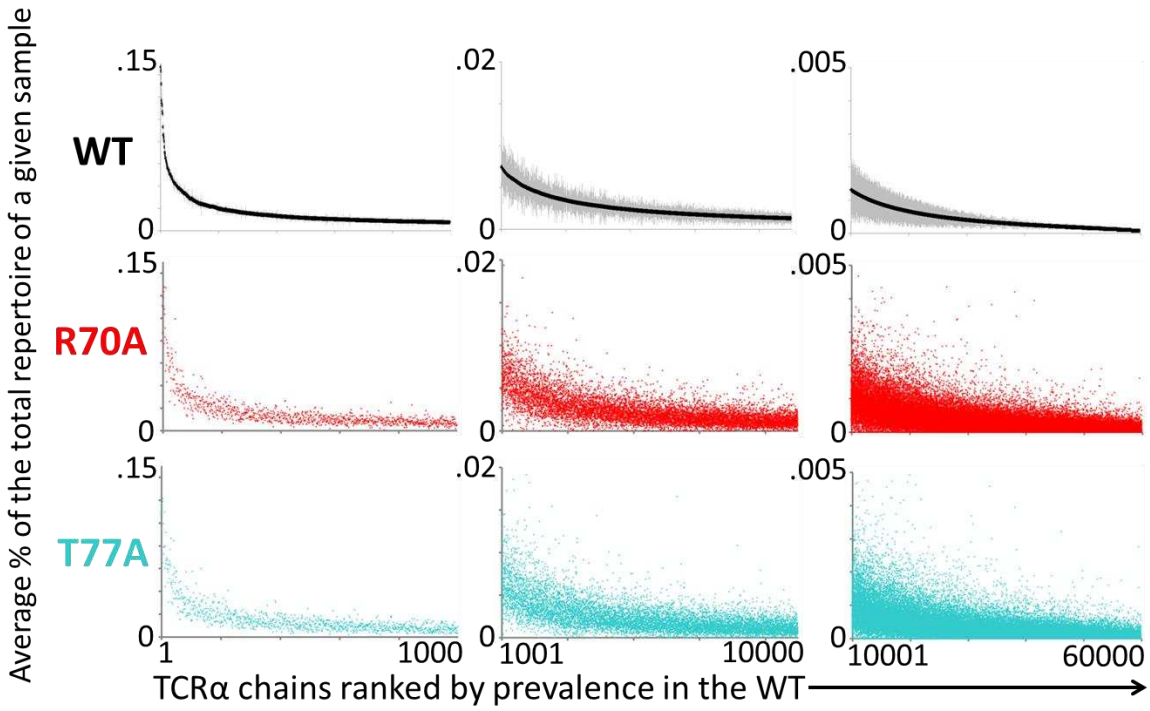

Figure 19: TCRα Ranked by Prevalence in WT Samples
All the TCRα chains found in all three WT replicates were ordered by their prevalence in the top panel. The position of these TCRαs in the mutants is show in the graphs below. While all 1000 top TCRαs are present in all three genotypes, there is some reduction in their abundance. In the second section, some TCRαs are completely missing from the mutants, and this number increases dramatically in the third section for the least frequent TCRα chains used. The WT graphs are indicated with black and the grey represent S.E.M. R70A is shown in red and T77A in teal, with S.E.M. omitted for visualization.

First, the sequence data from nine mice are represented as a 3-dimensional principle component analysis (PCA) (Figure 20). The WT, βT77A and βR70A mice clustered well and were separated from each other for two of the three components. Particularly well separated were the WT and βT77A data. As a second analysis the same software was used to do a direct comparison of the TRAV--CDR3-TRAJ combinations in the nine mice. Given the very large number of comparisons being made, the bar for significance differences was set very high. To reduce the number of comparisons a threshold of TCRαs sequenced at least 10 times in the data combined from all 9 runs was set, and, as shown in the heat maps in Figure 21, 84 combinations were found to be significantly different between WT and T77A mice. The figure gives a "gestalt" view of the data, while the complete data for these sequences, including the TRAV, TRAJ, CDR3 sequences and significance scores are contained in Table 11. In summary, these data show that despite the large diverse repertoires that develop in the WT and mutant mice, significant changes of TRAV family and subfamily as well as TCRα CDR3 sequences have occurred to accommodate the mutations. Furthermore, most of the TCRαs come from TRAV subfamilies that were differentially expressed between the T77A and WT shown in Figure 13, which reinforces the germline encoded bias of these TRAV subfamilies for interaction with MHC.

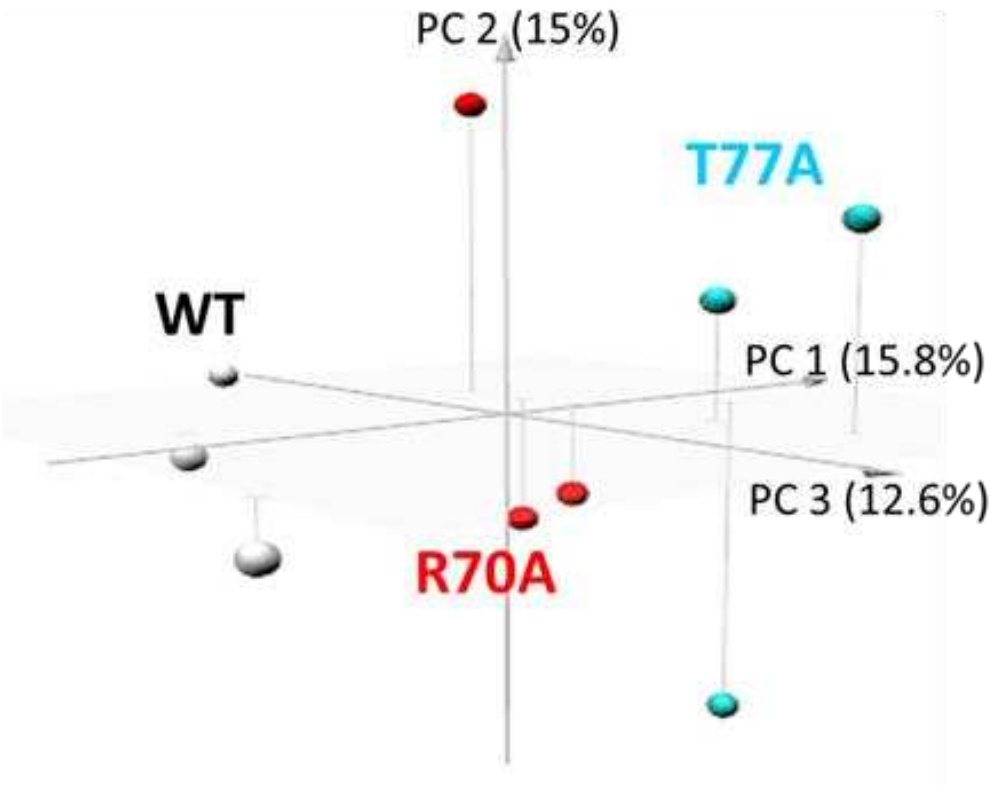

Figure 20: 3-D PCA of TCRα Repertoires
A principle component analysis of the nine TCRα sequenced repertoires was generated by DESeq2. This analysis suggests that there can be a grouping of each individual genotype upon deeper analysis. Particularly well separated are the WT and T77A. WT samples are shown with white balls while R70A are red and T77A are teal.

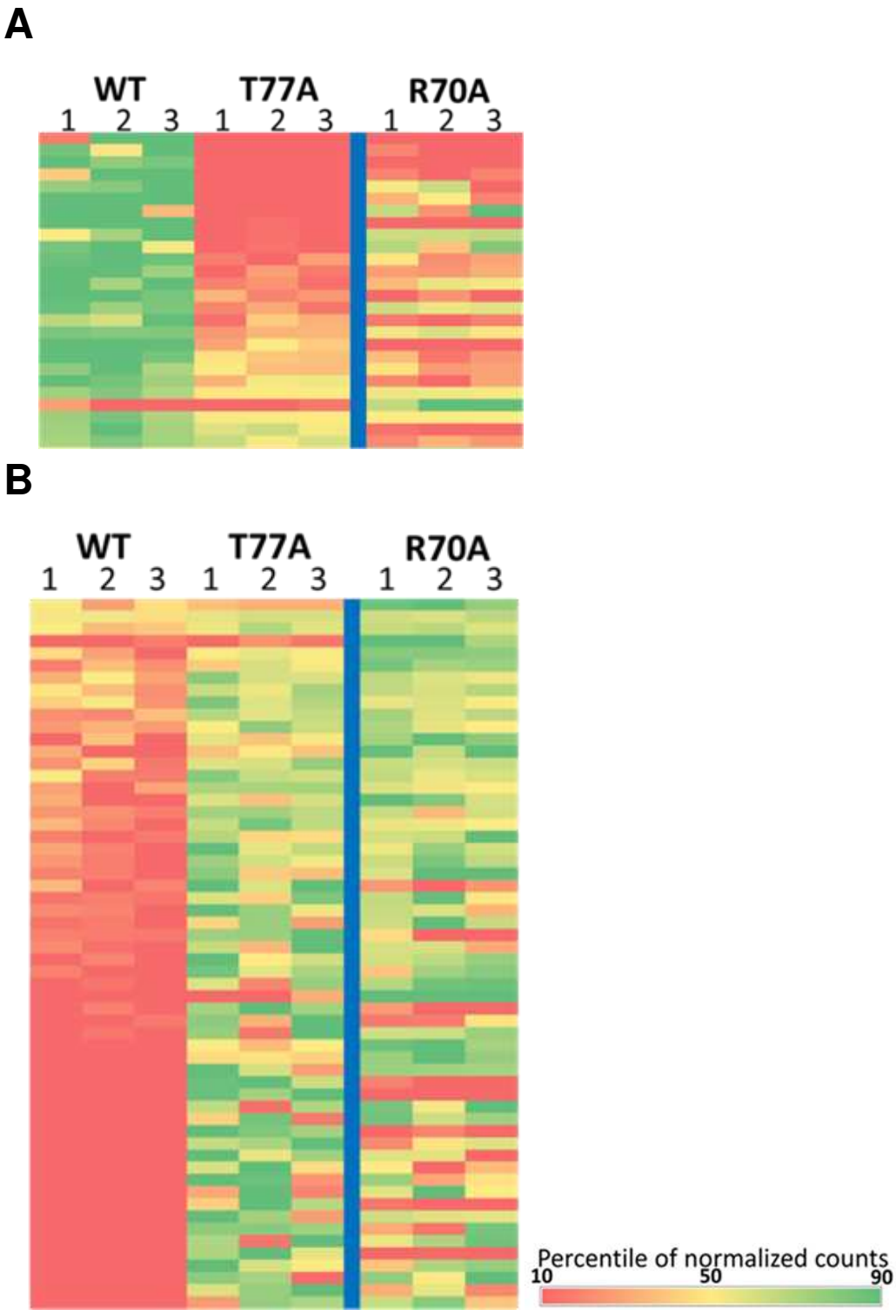

Figure 21: Differentially Expressed TCRα Chains

Heat maps of differentially expressed TCRα sequences between T77A and WT are shown. TCRα are ordered by positive (A) or negative (B) fold change. Green, yellow, and red indicate high, medium, and low usage, respectively. Statistical significance is indicated by a blue square for $p < .05$.

Table 11: Differentially Expressed TCRα Chains

TRAV subfamily, TRAJ, and CDR3	Log2 Fold Change	Adjusted P value
TRAV11 1N TRAJ04 CVVGAVLSGSFNKLTF	10.18	5.48E-03
TRAV11 1N TRAJ27 CVVAYNTNTGKLTF	9.81	7.88E-03
TRAV11 1AD TRAJ32 CVVVYYGSSGNKLIF	9.79	1.65E-06
TRAV01 1A TRAJ30 CAVSTNAYKVIF	9.79	7.29E-03
TRAV11 1AD TRAJ39 CVVGARGNNAGAKLTF	9.61	3.20E-02
TRAV01 1A TRAJ34 CAVRVPSNTNKVVF	9.55	7.56E-06
TRAV21 1A TRAJ56 CIMATGGNNKLTF	9.31	3.16E-02
TRAV11 1N TRAJ17 CVVGSAGNKLTF	7.27	1.37E-09
TRAV21 1A TRAJ56 CILVATGGNNKLTF	6.51	1.19E-02
TRAV21 1A TRAJ49 CILRADTGYQNFYF	5.84	3.38E-02
TRAV13 2AN TRAJ34 CAIDPNTNKVVF	3.93	2.17E-02
TRAV14 3A TRAJ15 CAASVGGRALIF	3.83	3.99E-03
TRAV11 1AD TRAJ57 CVVGVNQGGSAKLIF	3.82	9.71E-03
TRAV16 1N TRAJ21 CAMREDSNYNVLYF	3.10	5.25E-04
TRAV11 1AD TRAJ53 CVVGADSGGSNYKLTF	2.93	3.63E-02
TRAV11 1N TRAJ57 CVVGNQGGSAKLIF	2.74	4.21E-02
TRAV16 3D TRAJ45 CAMREGNTEGADRLTF	2.08	3.61E-02
TRAV01 1A TRAJ30 CAVRYTNAYKVIF	1.86	1.17E-05
TRAV14 2D TRAJ40 CAADTGNYKYVF	1.78	6.93E-03
TRAV11 1AD TRAJ18 CVVGSDRGSALGRLHF	1.60	1.96E-02
TRAV01 1A TRAJ28 CAVRPGTGSNRLTF	1.22	5.80E-03
TRAV19 1A TRAJ42 CAAGGGSNAKLTF	0.86	2.03E-02
TRAV11 1AD TRAJ21 CVVGPMSNYNVLYF	0.84	5.25E-04
TRAV06 5A TRAJ18 CALRRGSALGRLHF	0.79	1.18E-03
TRAV21 1A TRAJ47 CILRNYANKMIF	0.30	2.38E-05
TRAV21 1A TRAJ45 CILRVGAEGADRLTF	0.16	1.10E-02
TRAV19 1A TRAJ39 CAAGGNNNAGAKLTF	-0.17	8.53E-03
TRAV06 5A TRAJ34 CALSSNTNKVVF	-1.34	8.29E-03
TRAV11 1AD TRAJ42 CVVGPNSGGSNAKLTF	-1.94	4.96E-02
TRAV11 1AD TRAJ33 CVVGVSNYQLIW	-2.09	4.90E-04
TRAV03 3DN TRAJ37 CAVVTGNTGKLIF	-2.39	3.21E-02
TRAV11 1N TRAJ26 CVVGGNNYAQGLTF	-2.45	1.07E-02
TRAV06 7DN TRAJ27 CALGDRTNTGKLTF	-2.48	3.75E-02
TRAV03 3DN TRAJ27 CAVSASTNTGKLTF	-2.48	1.32E-02
TRAV10 1DN TRAJ23 CAARYNQGKLIF	-2.53	1.60E-02
TRAV03 1A TRAJ27 CAVSDNTNTGKLTF	-2.94	4.65E-02
TRAV03 3DN TRAJ13 CAVRANSGTYQRF	-3.01	3.84E-02
TRAV01 1A TRAJ09 CAVRDLGYKLTF	-3.07	3.59E-02
TRAV21 1A TRAJ57 CILRVPMNQGGSAKLIF	-3.16	1.88E-02
TRAV10 1A TRAJ21 CAASVSNYNVLYF	-3.19	3.25E-02
TRAV03 3A TRAJ32 CAVRGGSSGNKLIF	-3.27	3.61E-02
TRAV06 5A TRAJ44 CALSDLTGSGGKLTL	-3.85	1.60E-02
TRAV03 4A TRAJ39 CAVRNNAGAKLTF	-3.93	2.88E-02
TRAV03 3A TRAJ27 CAVSASTNTGKLTF	-4.09	8.11E-04
TRAV10 1A TRAJ37 CAVITGNTGKLIF	-4.18	4.61E-02

Table 11: Differentially Expressed TCRα Chains Cont'd

TRAV01 1A TRAJ18 CAVREGGSALGRLHF	-4.45	3.18E-02
TRAV10 1DN TRAJ12 CAARAGGYKVVF	-4.46	3.15E-02
TRAV03 3DN TRAJ27 CAVSGNTNTGKLTF	-4.47	5.80E-03
TRAV06 5A TRAJ53 CALSASGGSNYKLTF	-4.48	4.51E-02
TRAV21 1A TRAJ58 CILRVHGTGSKLSF	-4.58	7.62E-03
TRAV11 1AD TRAJ27 CVVGRNTNTGKLTF	-4.68	3.14E-03
TRAV03 1A TRAJ05 CAVSGTQVVGQLTF	-5.06	3.16E-02
TRAV11 1N TRAJ11 CVVGEDSGYNKLTF	-5.12	1.65E-02
TRAV11 1N TRAJ56 CVVATGGNNKLTF	-5.27	7.29E-03
TRAV11 1AD TRAJ18 CVVGAEGSALGRLHF	-5.55	2.48E-02
TRAV03 3DN TRAJ09 CAVRRNMGYKLTF	-6.02	1.60E-02
TRAV03 3A TRAJ40 CAVSARTGNYKYVF	-6.32	7.62E-03
TRAV11 1AD TRAJ26 CVVNYAQGLTF	-6.40	6.27E-03
TRAV11 1AD TRAJ40 CVVGAGNYKYVF	-6.42	1.84E-03
TRAV11 1N TRAJ17 CVVGARSAGNKLTF	-6.67	6.85E-04
TRAV11 1AD TRAJ06 CVVGAGGGNYKPTF	-6.73	2.28E-02
TRAV11 1AD TRAJ15 CVVGAKGGRALIF	-6.88	4.20E-02
TRAV01 1A TRAJ27 CAVTTNTGKLTF	-8.09	1.85E-06
TRAV01 1A TRAJ06 CAVTSGGNYKPTF	-8.11	3.22E-06
TRAV11 1AD TRAJ18 CVVVYRGSALGRLHF	-9.13	6.20E-03
TRAV01 1A TRAJ09 CAVRAMGYKLTF	-9.40	1.75E-04
TRAV11 1AD TRAJ27 CVVGAPGTNTGKLTF	-9.52	1.65E-06
TRAV11 1AD TRAJ32 CVVGEDYGSSGNKLIF	-9.53	2.07E-02
TRAV11 1AD TRAJ34 CVVGATSNTNKVVF	-9.57	2.07E-02
TRAV11 1N TRAJ06 CVVGLLTSGGNYKPTF	-9.67	6.12E-06
TRAV21 1A TRAJ31 CILRVAGNNRIFF	-9.71	5.25E-04
TRAV11 1AD TRAJ26 CVVGDNNYAQGLTF	-9.79	1.57E-02
TRAV11 1AD TRAJ38 CVVPNVGDNSKLIW	-9.92	2.88E-02
TRAV11 1AD TRAJ09 CVVVNMGYKLTF	-9.92	2.07E-02
TRAV11 1AD TRAJ39 CVVGAYNNAGAKLTF	-9.98	3.59E-02
TRAV11 1AD TRAJ27 CVVVHNTNTGKLTF	-10.03	4.90E-04
TRAV11 1AD TRAJ27 CVVGAENTNTGKLTF	-10.13	3.19E-03
TRAV21 1A TRAJ47 CILRVARDYANKMIF	-10.18	1.22E-02
TRAV11 1AD TRAJ27 CVVGAKDTNTGKLTF	-10.20	1.31E-02
TRAV11 1AD TRAJ26 CVVPYAQGLTF	-10.36	1.65E-06
TRAV11 1AD TRAJ06 CVVGLLTSGGNYKPTF	-10.57	5.90E-04
TRAV11 1AD TRAJ15 CVVGASQGGRALIF	-10.66	1.55E-02
TRAV11 1AD TRAJ31 CVVGVNSNNRIFF	-10.76	7.56E-06
TRAV11 1AD TRAJ27 CVVGDTGKLTF	-11.13	1.89E-04

The DESeq2 package calculated fold change and adjusted p values for differentially expressed TCRα between T77A and WT are shown in this table.

TCRβ Sequencing from αA64Q Mice

Now that I had seen a shift in the repertoire from the T77A mutant mouse, I decided to look deeper in the A64Q mutant for a similar shift. My analyses of the thymus in the αA64Q mice showed that there appeared to be a reduced activation from positive selection in DP thymocytes based on CD5/CD69 expression. Since substantial biological and structural data have shown A64 lies in a site that is often used for docking of βY48 in the CDR2 loop of the TRBV13~2 Vβ element and perhaps also the TRBV13~3 Vβ element (66), I focused my analyses on the effects of this mutation on the repertoire of T cells using these elements. I analyzed CD4 SP thymocytes and splenic $CD4^+$ T cells from WT vs. αA64Q mice with a mAb that discriminates TRBV13~2 from TRBV13~3 (Figure 22). Flow cytometric data showed that there was a substantial, significant shift in usage from TRBV13~2 to TRBV13~3 in both populations in the A64Q mutant versus WT mice.

Next, with a strategy similar to that used in my analyses of the TCRα repertoire in the βT77A and βR70A mice, I deep sequenced the TRBV13 domains present in naïve $CD4^+$ T cells in three WT and αA64Q mice. I created a PCR fragment with a 5'-primer common to all 3 members if the TRBV13 family and a 3'-primer within Cβ. Figure 23 shows that the sequence data confirmed the significant shift from TRBV13~2 to TRBV13~3 in the αA64Q mice, while there was no change in the use of the third family member (TRBV13~1).

Taking what I learned from my TCRα sequencing data, I went straight to DESeq2 for a deeper statistical analyses of these data. Looking at the PCA, even with some variability in the WT samples, I can still distinguish between the

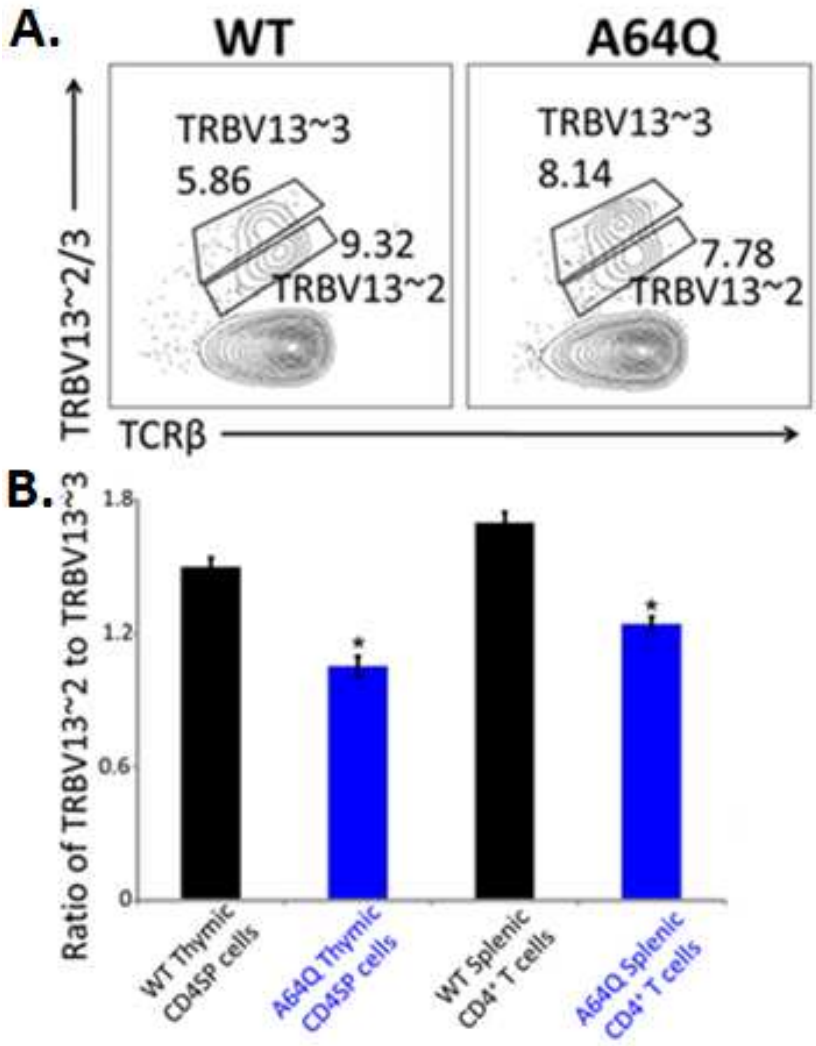

Figure 22: TRBV13 Subfamily Usage in the A64Q Mutant Mice
A.) Representative staining with mAb MR5-2, highlighting the two TRBV subfamily members it stains, on single positive CD4 T cells in the thymus is shown. B.) The ratio of these two subfamily members quantified over multiple mice and experiments, WT indicated with black bars and A64Q mice with blue bars, with error bars indicating S.E.M. and an * indicating $p<.05$ in a Student's t test.

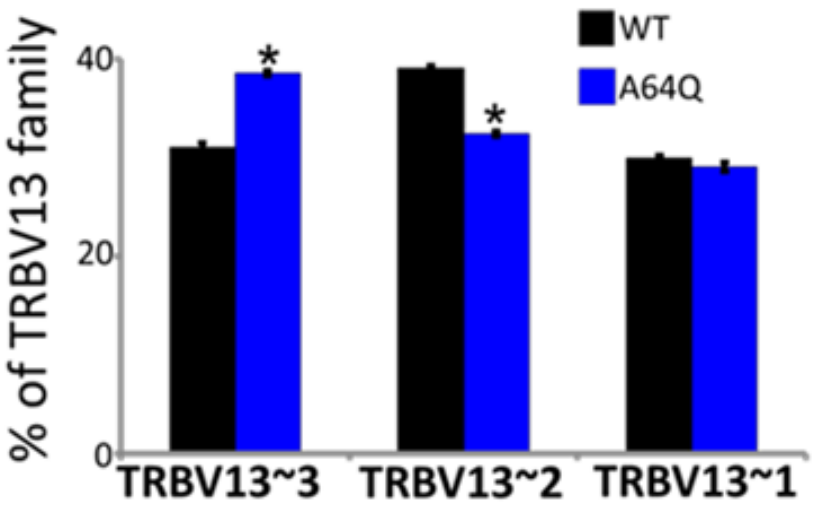

Figure 23: Subfamily Usage in TCRβ Sequencing
The different subfamily usage in the sequencing runs between WT and A64Q is presented in this graph with black bars for WT mice and blue bars for A64Q. This result recapitulates the staining data in which TRBV13~3 is used more in the mutant, TRBV13~2 is used less, and TRBV13~1 remains unchanged. The data are averages of three separate mice per group, error bars indicate S.E.M., and an * indicating $p<.05$ in a Student's t test.

genotypes (Figure 24). The top section of the heat map in Figure 25 shows all the TRBV13/TRBJ combinations with an increased frequency in WT samples compared to A64Q while the bottom section shows the combinations used more frequently in the mutant. The blue squares in Figures 25 indicate statistical significance (Table 12), and these combinations group almost perfectly with TRBV13 subfamily. Furthermore, looking at the more commonly found TCRs, DESeq2 identifies individual TCRβs that are significantly differentially expressed between the WT and A64Q (Figure 26, Table 13). Thus, these data support on a more global scale the previous findings with particular TRBV13~2 containing domains associating the TRBV13~2 CDR2 loop with docking on the portion of the MHCII β1 helix containing with an evolutionary preference for MHCαA64.

Functional Differences of the TCR Repertoires in the Mutant Mice

While the differential expression of different TCR V families and subfamilies clearly highlights the evolutionarily conservation of the TCR-MHC interaction, the subtlety of the effect raises questions about the biological relevance of such an effect. To begin to address this question, I performed an experiment to assess how "foreign" the WT and mutant I-A^b molecules appeared to $CD4^+$ T cells from the various mice. I set up one-way mixed lymphocyte reactions using all combinations of purified $CD4^+$ T cells and antigen presenting cells (APCs) from the WT and mutant mice. T cells and APCs from an I-A^f mouse (B10.M) were used as a control (Figure 27). As these are true alloreactive-like assays, with the only difference in these mice being a single point mutation on the MHCII and no superantigen or minor antigen differences, I expected only a small percentage of the T cells to actually

Figure 24: PCA of Replicate WT and A64Q TCRβ Sequencing

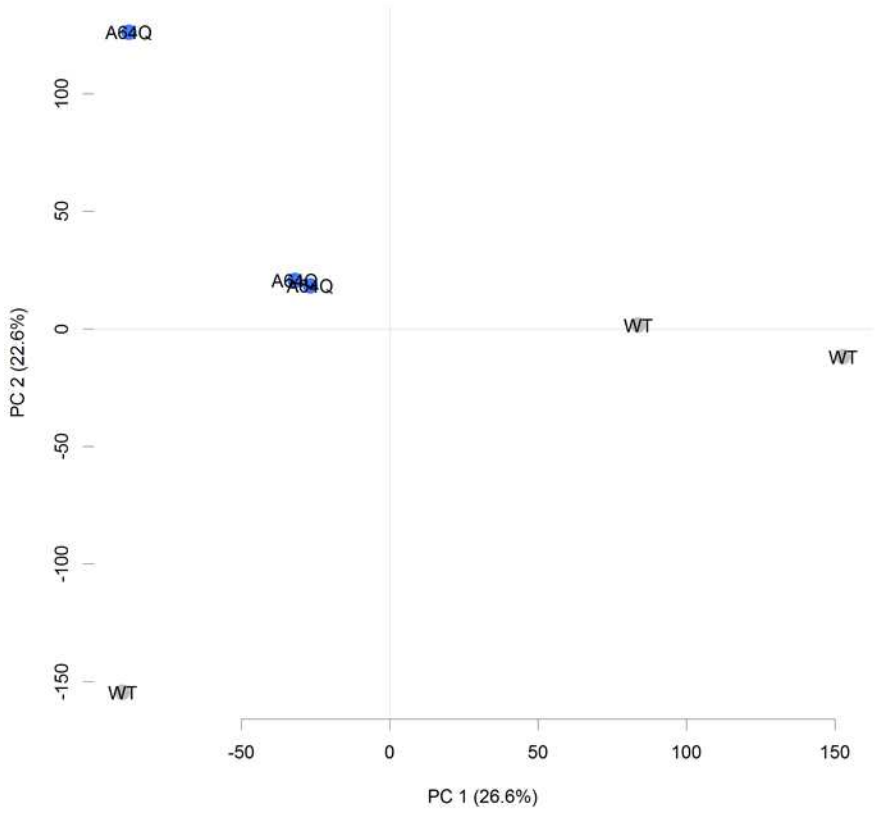

Figure 24 shows a PCA of the TRBV13 sequencing on A64Q and WT samples. While there is more variability in some of the WT samples, this plot shows genotypes are distinguishable from each other.

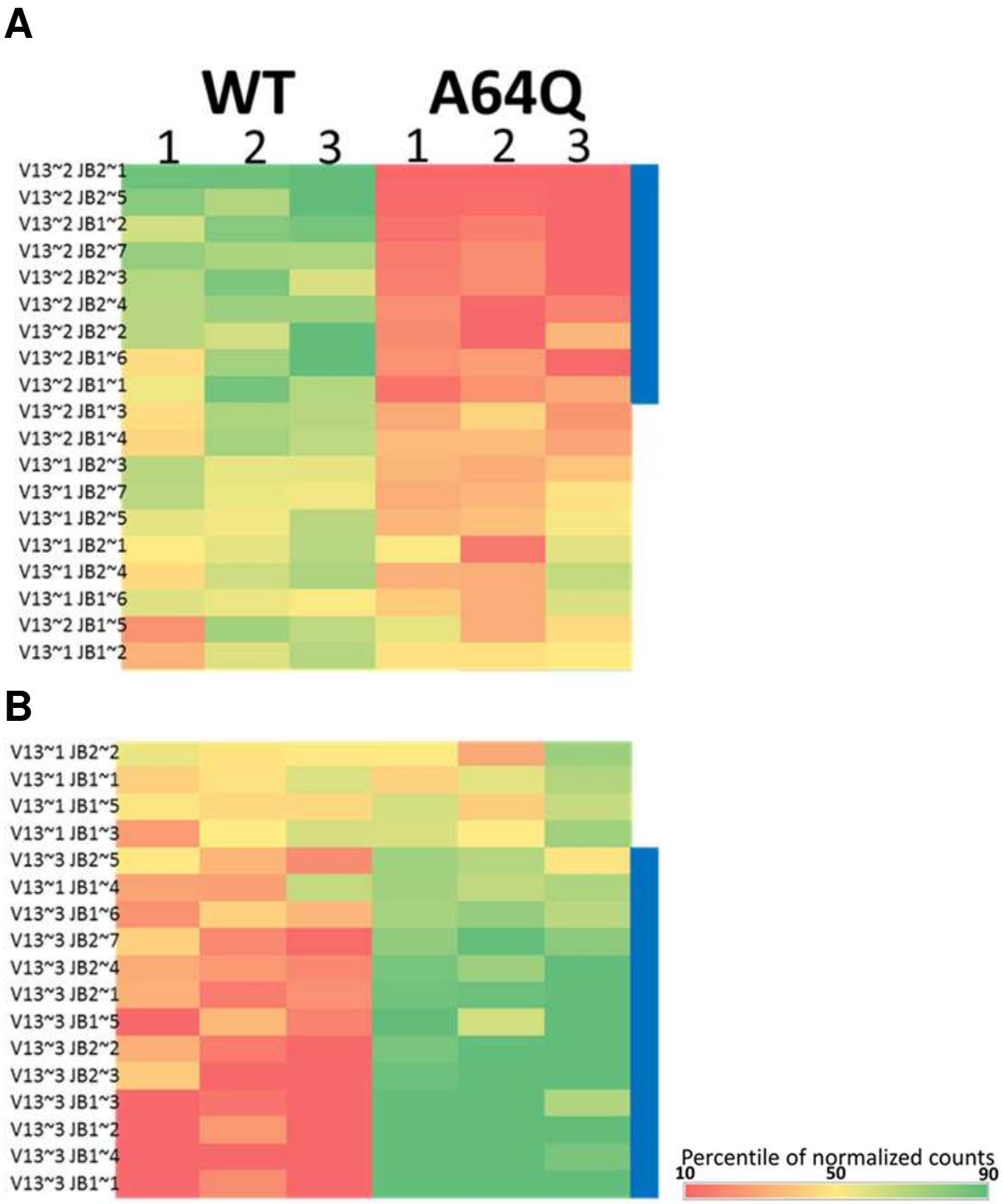

Figure 25: Heat Map for Subfamily and TRBJ Combinations in A64Q Mutant vs. WT Mice

Heat maps of TRBV13-TRBJ combinations in the sequencing runs between the A64Q mutant and WT are presented. The heat map ordered by positive (A) or negative (B) fold change. Green, yellow, and red indicate high, medium, and low usage, respectively. Statistical significance is indicated by a blue square for $p < .05$.

Table 12 DESeq2 Statistics for Subfamily and TRBJ Combinations in A64Q Mutant vs. WT Mice

TRBV-TRBJ	Log2 Fold Change	Adjusted p value
V13~2 JB2~1	0.41	3.33E-15
V13~2 JB2~5	0.31	8.30E-09
V13~2 JB1~2	0.26	9.72E-07
V13~2 JB2~7	0.25	8.63E-06
V13~2 JB2~3	0.23	4.15E-05
V13~2 JB2~4	0.23	5.51E-07
V13~2 JB2~2	0.22	1.38E-05
V13~2 JB1~6	0.21	6.68E-04
V13~2 JB1~1	0.19	1.88E-04
V13~2 JB1~3	0.10	7.50E-02
V13~2 JB1~4	0.10	7.47E-02
V13~1 JB2~3	0.09	7.47E-02
V13~1 JB2~7	0.07	1.51E-01
V13~1 JB2~5	0.06	2.43E-01
V13~1 JB2~1	0.06	3.02E-01
V13~1 JB2~4	0.05	3.02E-01
V13~1 JB1~6	0.03	5.67E-01
V13~2 JB1~5	0.02	7.40E-01
V13~1 JB1~2	0.00	9.99E-01
V13~1 JB2~2	-0.02	7.26E-01
V13~1 JB1~1	-0.05	3.25E-01
V13~1 JB1~5	-0.06	3.02E-01
V13~1 JB1~3	-0.09	1.02E-01
V13~3 JB2~5	-0.15	2.24E-02
V13~1 JB1~4	-0.15	7.17E-03
V13~3 JB1~6	-0.19	3.34E-04
V13~3 JB2~7	-0.27	6.12E-06
V13~3 JB2~4	-0.28	4.39E-09
V13~3 JB2~1	-0.30	5.07E-09
V13~3 JB1~5	-0.33	1.86E-08
V13~3 JB2~2	-0.35	1.37E-09
V13~3 JB2~3	-0.40	4.88E-09
V13~3 JB1~3	-0.42	1.38E-10
V13~3 JB1~2	-0.48	4.83E-13
V13~3 JB1~4	-0.48	6.96E-15
V13~3 JB1~1	-0.51	7.95E-20

The statistical output of DESeq2 for the A64Q to WT comparison of TRBV13-TRBJ combinations found amongst the sequencing runs is shown in this table.

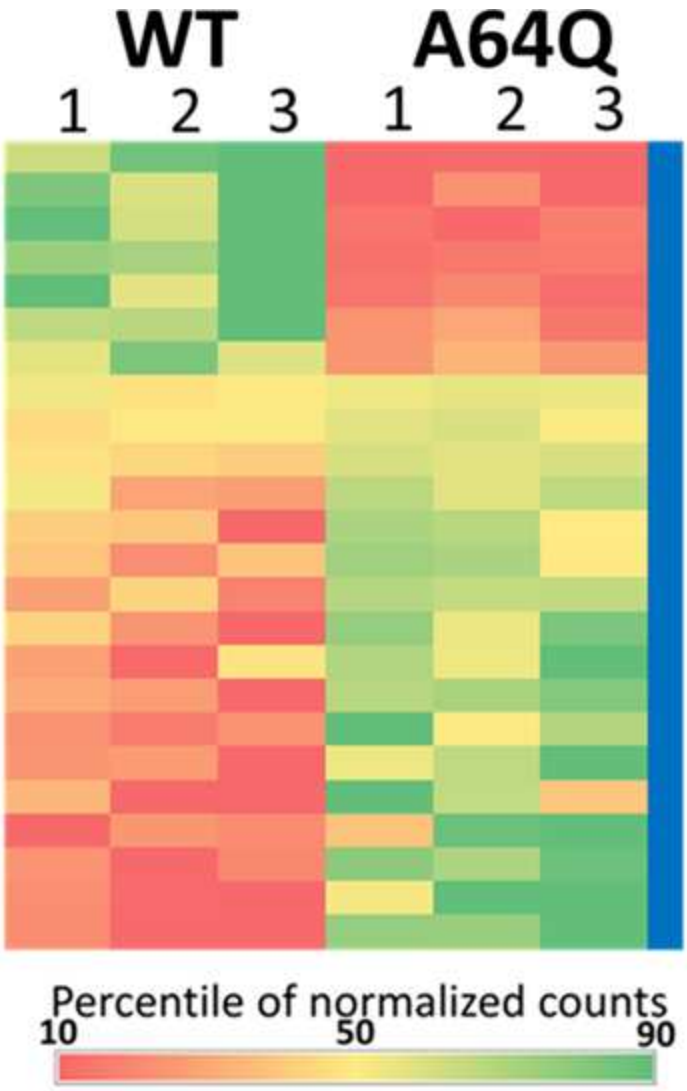

Figure 26: Differentially Expressed Individual TCRβs Between A64Q Mutant and WT Mice

The heat map for individual TCRβs statistically significantly differentially expressed between A64Q mutant and WT mice as found by the DESeq2 package is presented. Red indicates lower usage in those samples. These results suggest a repertoire of TCRβs favored by each genotype.

Table 13: DESeq2 Statistics for Differentially Expressed Individual TCRβs Between A64Q Mutant and WT Mice

TRBV13-TRBJ-CDR3	Log2 Fold Change	Adjusted p value
V13~2 JB2~5 ASGDDRGQDTQY	3.94	4.49E-02
V13~2 JB1~1 ASGDGTANTEVF	3.27	4.49E-02
V13~1 JB2~4 ASSDPGQNTLY	2.99	3.31E-02
V13~1 JB2~7 ASSDAGGTYEQY	2.78	2.40E-02
V13~2 JB2~7 ASGDARGSYEQY	2.58	4.80E-02
V13~2 JB2~5 ASGDGGNQDTQY	1.66	4.42E-02
V13~1 JB2~3 ASSDSAETLY	1.14	4.16E-02
V13~3 JB2~3 ASSDSAETLY	-0.82	4.42E-02
V13~3 JB1~3 ASSDRDSGNTLY	-0.96	4.42E-02
V13~2 JB1~1 ASGDAGQNTEVF	-1.24	7.84E-03
V13~3 JB2~7 ASSDAGYEQY	-1.68	4.49E-02
V13~3 JB2~3 ASSAETLY	-1.87	4.49E-02
V13~3 JB2~3 ASSDPGSAETLY	-2.02	4.16E-02
V13~3 JB2~5 ASSDSQDTQY	-2.05	4.49E-02
V13~1 JB2~7 ASSDALGSSYEQY	-2.49	4.16E-02
V13~1 JB1~1 ASSEQANTEVF	-2.75	4.49E-02
V13~3 JB2~2 ASSENTGQLY	-2.78	2.40E-02
V13~3 JB2~5 ASSDWDTQY	-3.01	2.40E-02
V13~3 JB2~3 ASSDRGTSAETLY	-3.18	4.16E-02
V13~3 JB2~4 ASSDESQNTLY	-3.39	4.16E-02
V13~3 JB2~1 ASSEGTGGNYAEQF	-3.59	4.49E-02
V13~3 JB1~3 ASSGQSGNTLY	-3.80	2.45E-04
V13~3 JB2~2 ASSDGTANTGQLY	-4.12	5.91E-03
V13~3 JB2~3 ASSETGGSAETLY	-4.39	2.76E-02

The statistical output of DESeq2 for the A64Q to WT comparison of individual TRBV13-TRBJ-CDR3 found amongst the sequencing runs is shown in this table.

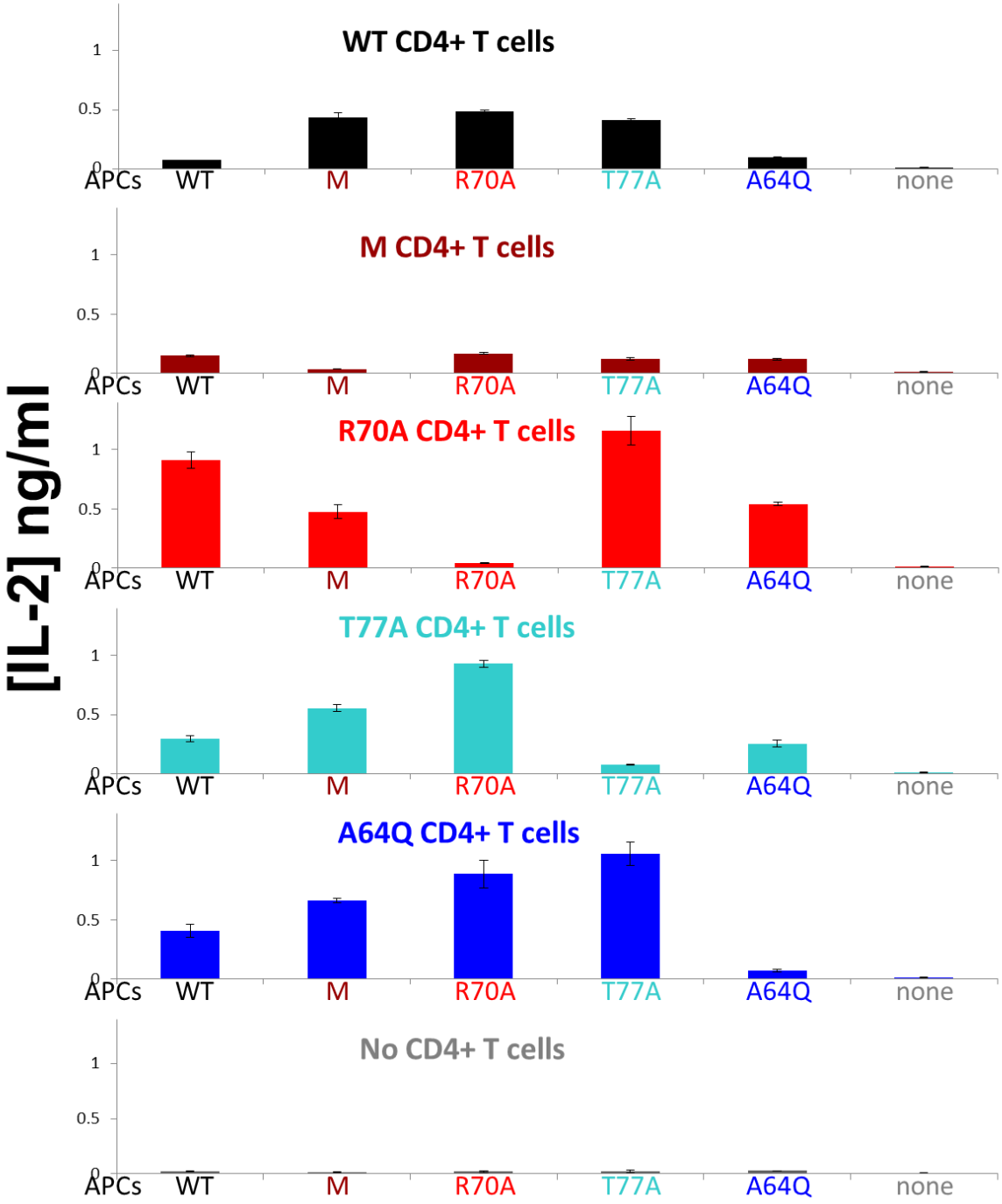

Figure 27: Functional Differences of the Mutant TCR Repertoires
One-way mixed lymphocyte reactions using all combinations of purified $CD4^+$ T cells and APCs from the WT and mutant mice were conducted. T cells and APCs from an $H\text{-}2^f$ haplotype mouse (B10.M) were used as a control. Data shown are IL-2 production assays from 3 independent experiments with error bars indicating S.E.M.

respond. Because of the small population of cells actually responding, I measured IL-2 in the cultures are an indication of response at 4-6 days. IL-2 is a good indicator for response, but is not linearly quantitative for the initial response, as responding cells both make and consume the cytokine. The results show that the $CD4^+$ T cells did not respond to APCs from the same mouse, but responded to APCs from all of the other mice. The magnitudes of the responses seen upon the WT and mutant I-A^b mice were on the same order of magnitude as the alloreactive responses seen with the I-A^f T cells and APCs. These results predict that differences in the TCR repertoires among the WT and mutant I-A^b mice should be similar to that among mice of different MHCII haplotypes indicating TCR repertoire is different enough to be functionally relevant.

Discussion

In this chapter I addressed the heart of how the mutations directly affected the use of germline encoded V regions and highlighted their evolutionary bias for MHC. I started with the βT77A mutant, using the βR70A as a control, and explored the TRAV usage of CD4 T cells in the thymus and periphery of these mutants compared to WT mice. Noticeably, with one of the antibodies for staining, there was a reduction in the TRAV14 usage in T77A mice. I then sorted naïve CD4 T cells from the spleen of all the mice and sequenced their TCRα chains. I found on the TRAV family level no difference in usage between WT and R70A mice, but there were statistical differences between WT and T77A. Furthermore, I observed differences in subfamily usage between WT and T77A, most notably in the TRAV14 3A

subfamily, which corroborates the staining data seen previously. These data are the clear evidence of an evolutionary bias of some TRAV for recognition of the MHC.

Next I examined the repertoires in the mutant mice for their individual TCRα, including randomly recombined CDR3 region. The mutant repertoires have similar richness and diversity when compared to WT, and the sequencing has not saturated the population. However, the repertoires are not random dispersed highlighting a biological significance in the sequences found repeatedly. After trying several different comparisons of the mutants to WT samples, ultimately the data were compared using DESeq2, a package designed for RNAseq data. When looking at the TCRαs found repeatedly, this package was able to identify differentially expressed TCRαs between all the samples. The TCRαs are reduced in the mutants highlight the effect of the mutations, while the ones that are increased suggest how the mutant repertoires are compensating.

Next, having already identified that the thymocytes developing in the αA64Q mice were receiving less of a signal during positive selection, I sought out to determine if their repertoire of TRBV subfamilies was also shifted. Again, I accomplished this by using antibody staining and flow cytometry to identify TRBV13~2 and TRBV13~3, the former decreased in the mutant and the latter increased. I confirmed this result by sequencing TCRβ chains from naïve CD4 T cells in these animals. When looking at TRBV-TRBJ combinations that are differentially expressed between the mutant and WT samples, they group almost perfectly by TRBV subfamily, highlighting the mutant effect on the TRBV13~2 subfamily and the evolutionary bias of this subfamily towards MHC. Once again,

DESeq2 can identify differentially expressed TCRβ decreased in the mutant, highlighting the effect of the mutation, and increased TCRβ in the mutant, suggesting the compensation in the repertoire.

Finally, I sought a functional measure of the subtle differences in the repertoires. To accomplish this, I performed one-way alloreactivity assays with CD4 T cells from the WT and mutants responding against APCs from the other genotypes. Encouragingly, the T cells from every genotype did not respond to their own APCs, but responded to all other genotypes, highlighting the functional differences of the mutant and WT repertoires.

Although the TRAV primers were designed to be family specific, of the similar length and of similar melting temperatures, I only expected my results to be quantitative in comparisons within TRAV families, but just semi-quantitative between TRAV families. Nevertheless, the biases in analysis between each TRAV family should be common to the different types of mice, so I felt comparisons of TRAV usage between mouse genotypes were justified.

It is important to note that all the TCRα and TCRβ sequenced in these experiments can have variable partners on the other chain, so two TCRα chains sequenced in both the WT and T77A might actually be paired with different TCRβ chains that may possess varying biases towards MHC. Furthermore, based on the number of out of frame sequences, it is evident that I am sequencing the non-selected chain as well at some frequency. For these reasons, I am probably underestimating the differences in the repertoires of my mutant mice using the sequencing data. Even if I was to sequence paired αβ TCRs from the same cell, it is

likely that the repertoires in the mice are so large that sampling of such a huge population could become problematic for my analyses.

In addition to looking at TRAV usage, I could look at TRAJ gene usage either overall or for specific TRAVs. There is a prevailing idea in TCRα rearrangement that the presence of distal TRAJ elements is a sign of secondary rearrangements occurring (117, 118), which could be prompted by a T cell's failure to be positively selected with its initial rearrangement. However, it is also known that certain TRAV genes are preferentially rearranged with certain TRAJ genes based on chromosomal location (117, 118). Thus, I must be careful in my interpretation of these data because as I see a shift in TRAV usage that alone could account for a shift in TRAJ usage. Even if I look at TRAJ usage on an individual TRAV basis, that does not give me evidence of secondary rearrangement to that TRAV, as it is impossible to rearrange to the same TRAV on the same chromosome in the same cell more than once (119). Therefore, even if there is a slightly different TRAJ usage that might be a sign of secondary rearrangements following failed selection on certain TRAVs, it is impossible in my system to discern which of these genes are being most selected against, so I did not focus my analyses on these elements.

CHAPTER VI

DISCUSSION AND FUTURE DIRECTIONS

Discussion

The roots of my current thinking on the evolutionary conservation of interactions between TCR and MHC amino acids came from studies on the $CD4^+$ T cells that develop in mice expressing a single fixed peptide-MHCII complex (62). These mice had impaired negative selection due to the absence of a diverse set of self-peptides bound to their MHCII, yet they still developed a very large population of peripheral $CD4^+$ T cells. This result gave our lab insight into the repertoire of T cells that exist in the thymus after positive selection and before negative selection. They demonstrated that these T cells not only responded with high frequency to self-MHC occupied by the normal complement of self-peptides, but also, surprisingly, to many different allo-MHCII alleles. Most of these cells disappeared when bone marrow expressing normal MHCII was introduced into these mice. Our lab concluded that although negative selection functions to remove high affinity self-specific T cells, in so doing, it also eliminates a large population of highly MHCII cross reactive T cells.

Following up on this idea, our lab tested the role of the CDR3 in one of the TCRs that had not been subjected to normal negative selection (82). In this study, mutagenesis was conducted to just the CDR3β of one TCR that was both peptide specific and alloreactive to several MHCs. The study indicated that different CDR3β sequences, completely unrelated from the original CDR3β sequence, were able to alter independently both the specificity or alloreactivity, even creating new alloreactivity to different MHCs, but all were still dependent on germline encoded

interactions. Thus, for this given TCR, the somatically rearranged CDR3β was able to tune the peptide as well as the MHC specificity on several different haplotypes.

Based on the subsequent functional, mutational and structural studies with these crossreactive T cells, our lab concluded that the TCRs' high cross reactivity was due to the dominant interaction of certain conserved amino acids in their TRAV and TRBV CDR1 or CDR2 loops with conversed sites on the helieces of various MHCII alleles. In complexes between the TCRs of T cells from normal mice and their antigenic peptide-MHCII ligands, reported by ourselves (63) and others (64), these conserved interactions were also often seen, but they were not usually so dominant. These findings have led me to my current hypothesis that random combinations of germline TCR α and β genes create T cells reactive to MHCII regardless of allele with high frequency. Yet, in order to escape negative selection and contribute to the functional peripheral repertoire, T cells must bear TCRs whose somatically generated CDR3s have modulated this tendency away from generic MHC reactivity and toward peptide dependence.

The results of other experiments by our lab and others are consistent with this idea. One study was conducted using mice deficient in terminal deoxynucleotidyl transferase (TdT) expression (120). These mice lack the enzyme responsible for non-germline encoded junctional nucleotide additions and thus bear TCRs that rely exclusively on the germline nucleotides to generate their CDR3 regions. These TCRs consistently show increased affinity for pMHC but reduced peptide specificity. The CDR3 tuning hypothesis explains this TCR promiscuity as coinciding with accentuated affinity for the α helices of the MHC. While these are examples of

functional tuning for some TCRs, there have also been structural (63, 121) and biochemical analyses (122). Crystal structures have visualized the changing of germline encode contacts to MHC due to CDR3 regions and biochemical mutational experiments explain how a given CDR can tune a TCR-pMHC interaction. These results, taken together with the mutational data disrupting the germline encoded interactions (66), set the table for this study, in which I was able to assess the tuning of CDR3s in a flexible system without fixed TCR chains.

My purpose in this present study was to determine how mutations in MHCII I-A^b amino acids affect T cell development and the peripheral T cell TCR repertoire. I chose I-A^b βT77 and αA64 for this study for several reasons; they are highly conserved among MHCII molecules, they have been reported repeatedly as sites of interaction with certain germline TRAV CDR1 and TRBV CDR2 amino acids, their mutations often disrupts the activation of peripheral antigen specific T cells in response to antigen, and they do not participate directly in peptide binding. I choose I-A^b Rβ70 as a control because it is less conserved and usually interacts with the TCR CDR3 loops, so the TCRs might be able to compensate by using different residues in their CDR3.

My results show that none of the mutations prevented the development of a large, diverse peripheral $CD4^+$ T cell population. However, depending on the mutation, there were significant changes in thymocyte subpopulations and changes in the peripheral $CD4^+$ T cell TCR repertoire. The subtlest changes were seen in mice harboring the βR70A and βT77A mutations. There were no changes with these mutations in either thymic cellularity or the proportion of the thymic population

undergoing selection ($CD5^+CD69^+$). However, compared to WT mice, the βT77A mutation led to significant shifts in TRAV family and subfamily usage. In addition, this mutation led to changes in the TRAV-CDR3-TRAJ repertoire, demonstrated by the fact that PCA clearly separated the unique sequences in the mutant mice from WT mice and from each other. My analyses of the αA64Q mice also showed normal thymic cellularity, but in this case there was a significant reduction in activation in thymocytes undergoing selection. In a more abbreviated peripheral repertoire analysis, I examined the usage of TRBV13~2, compared with the other two members of this family. The importance of the intimate interaction of the evolutionarily conserved Y48 in the CDR2 of TRBV13~2 with the region of MHCII α chain helix containing A64 has been documented in numerous structural, functional and thymic developmental studies (63, 66, 123). While still present, the importance of this amino acid in the other family members is not as clear. My analyses showed that TRBV13~2 usage by both thymic and peripheral $CD4^+$ T cells is reduced in the mutant mice with a concomitant rise in TRBV13~3 and no change TRBV13~1 compared to the WT mice.

The results of the present study clearly show that mutation of either βT77 or αA64 alters the repertoire of developing CD4 T cells. However, the magnitude of these effects was less than what I had previously observed in the response of antigen-primed WT peripheral $CD4^+$ T cells to antigenic peptides presented by the mutant MHCII proteins. Likewise, mutation of conserved amino acids in the CDR2 loop of TRBV13~2 had a much more profound effect on T cell development than did the αA64Q mutation (66). These results suggest to me that during the development

of the TCR repertoire, adjustments not only in V family usage, but likely also in αβ pairing and somatically generated CDR3 sequences can largely compensate for the loss of a single conserved docking site on MHCII. However, once a T cell has been selected by WT MHCII, it can no longer make these adjustments to the loss of the docking site. This idea is consistent with modelling and functional studies that explore the energetics for TCR-MHC interaction hot spots for crossreactive and monospecific TCRs (124). It is also worth noting that our lab's previous results with mutations in TRBV-13.2 CDR2 were done with a transgenic TCR β chain with a fixed CDR3, thus limiting the possible adjustments in repertoire to only changes in α chain pairing.

It has been suggested that the great deal of latitude seen in the docking angle of TCRs binding to MHC, argues against structural pairing of evolutionarily conserved amino acids in TCR-MHC interactions. However, the identification of the conserved interaction of TRBV-13.2 CDR2 with the MHCII α chain was first suggested by the many TCR-MHC structures that have been solved showing repeatedly that conserved amino acids in this CDR2 loop interact with the same site on the MHCII α1 helix even in the face of various docking angles of the TCRs. In the dozens of other TCR-MHC structures solved, there has not been such an extensive a set of structures available for analysis involving other TRAV and TRBV elements. However, since many of the residues in the MHC helices are conserved, it is possible that individual TRAV or TRBV elements prefer docking to different conserved sites or can use alternate ones to the preferred site.

A recent study consistent with this idea comes from the Garcia laboratory (125). They screened a peptide library and found a series of peptides that when bound to the MHCI allele, H2-L^d (L^d), can be recognized by the same TRBV13~1 containing TCR. They solved a series of structures of the TCR bound to the peptide-L^d ligands. The results show that, while the TRAV CDR1 and CDR2 locations on the L^d α2 helix were very similar in the structures, the TRBV13~1 CDR1 and CDR2 loops had more than one docking site on the L^d α1 helix, altering the angle of engagement of the TCR with L^d. Interestingly there were discrete docking positions, not a continuous series. These results establish multiple discrete conserved sites for TRBV13~1 docking on MHCI, the choice of which is determined by the peptide. Therefore, the single amino acid mutational approach used here may make it difficult to establish completely the TRAV or TRBV partners for a particular conserved site on the MHC helices.

The results of this study are not inconsistent with any of the recent studies that have shown highly unusual MHC docking modes by some TCRs and non-MHC ligands for some TCRs. For example, NKT and MAIT T cells have non-conventional MHC ligands that lack the conserved MHC docking sites (126, 127). They have specialized canonical Vα domains that pair with a subset of conventional Vβ domains. Their TCRs dock in very unconventional ways on their ligands. These specialized T cells and their ligands arose evolutionarily after the development the conventional TCR-MHC system. One could consider that they have “hijacked” a part of system for another purpose, much as certain MHC-like molecules no longer

function as ligands for T cells but have taken on new functions over evolutionary time (128).

The set of conventional TCRs that deviate most in the orientations and locations with which they interact with conventional peptide MHC complexes primarily come from autoreactive T cells. Their footprints on MHC can drift dramatically away from those seen with foreign peptide-MHC complexes, and in one case, even reverse the orientation of the TCR on the ligand (58). These T cells are the survivors of thymic negative selection and as such may need to venture into these unusual docking modes, not found in the thymus, to improve their affinity to achieve T cell activation by self-peptides.

The results of this study are also not inconsistent with previous studies that have argued against the evolutionary hypothesis. In these experiments TCRs were under great selective pressure to use CDR3s to allow them to recognize either MHC in an unusual manner (46, 58, 70, 129) or even non-MHC ligands (68, 130). The most dramatic example comes from the Singer laboratory (67-69). This laboratory constructed a mouse lacking MHCI, MHCII, CD4 and CD8. They also introduced mutations that uncoupled essential downstream TCR signaling molecules from essential interactions. The mice develop a peripheral T cell repertoire that contains T cells reactive to the surface protein, CD155. The authors conclude that these experiments show that the TCR repertoire need not be MHC dependent and that the usual specificity for MHC is not inherent in the germline sequences of the MHC and TRAV/TRBV elements. Rather, they suggest that in normal mice, MHC specificity arises by selection from a somatically generated random repertoire of TCRs, yielding

TCRs that can satisfy the MHC-dependent geometry of the many components of the large TCR/co-receptor signaling complex.

My experiments do not argue against the generation of T cells of these non-MHC specificities. In fact, given the recombinatorial capacity of the thymus to generate an enormous number of unique TRAV and TRBV CDR3 loops, their existence is inevitable. But, if the initial, unselected TCR repertoire is random, the frequency of T cells specific for any particular protein, such as CD155 or MHC proteins themselves will be very low, as it is in the case of antibodies. Subsequent culling of this scarce MHC-specific repertoire during the non-proliferative selective phase of T cell development to make it both self MHC-restricted and self-MHC-tolerant will further greatly reduce its size, making the generation of the well-established very large peripheral T cell repertoire very difficult. But predisposing the pre-selection TCR repertoire toward MHC recognition via embedded conserved amino acids in MHC and TCR proteins to promote their interaction, should separate the “wheat” from the “chaff” during selection much more efficiently, while prohibiting the generation of rare TCRs specific for other proteins.

Evidence presented in this thesis as well as previous papers suggests that this latter idea is to some extent correct, as the preselected TCR repertoire is already skewed towards reactivity (43, 51-53, 66, 81). However, it is important to keep in mind that the same logic applies to experiments on TCRs that have not undergone positive selection, yet still have a bias for MHC. These experiments that show that the preselected repertoire still has a bias for MHC are really just a factor of how CDR3s modify germline encoded interactions, and the readout of MHC

reactivity is based off of the peptide repertoire presented by a given MHC molecule. While answering what percent of a given population is able to still recognize certain MHCs, those experiments do not really address what are the important germline encoded interactions.

While the results of this study clearly highlight the evolutionarily conserved interactions of the TCR V subfamilies to the areas of the mutations, the subtle nature of these shifts highlights the flexibility in the system during thymic selection. The greater effect that these mutations have on mature T cell responses rather than during development yields an appreciation for the ability of TCRs to rearrange new CDR3s and make new αβ pairings that can still recognize a diverse repertoire of peptides presented by MHC molecules, as well as other factors involved in thymic selection not mediated by specific TCR-pMHC interactions.

Future Directions

Due to the subtle nature of the results of this study highlighting the germline encoded interactions of different subfamilies, this thesis could be the first step in teasing apart these interactions. If I wanted to observe a more drastic shift in the TCRα repertoire in the MHCβ mutant mice, fixing a TCRβ chain and making the mice TCRα$^{+/-}$ might be more revealing. Depending on the TCRβ chain chosen, this future direction could further highlight more drastically the germline encoded rules that govern interaction with the MHCβT77 area. From this study, I can predict the subfamilies that would be least able to compensate for the mutation, and help narrow the search for the structural basis behind these interactions.

Originally when I designed these experiments, I hypothesized that perhaps there would be large shifts in the numbers of T cells selected in the thymus. If such large shifts existed, then deeper analysis of the repertoire would have really been informative for specific interactions in the germline encoded portions of the TCR that govern MHC recognition. In addition to answering this age-old immunological question, understanding these rules might be the most translationally relevant aspect of this research. Since it is known that certain TRAVs are heavily biased in either the response towards specific pathogens, as well as in autoimmune responses, knowledge about the germline encoded interactions could be useful for modulating either of these responses as a therapeutic in people.

The flexibility of TCR-pMHC interaction highlighted in this thesis as well as in previous experiments (67), poses an interesting obstacle for future studies of germline encoded interactions. All studies on this topic hoped to make generalized rules about germline encoded TCR residues that interact with MHC, but to this point it has been really difficult to separate these rules from the interactions of randomly recombined CDR3 regions interacting with a multitude of peptides presented by different MHC molecules. Recent advances in technology for measuring TCRs interacting with MHC involving NMR and bioforce-probe analyses have elucidated interactions previously undetectable (131). In this study, the authors measure interactions of a TCRβ chain paired with pre-Tα and show its MHC specificity. While this TCRβ has a fixed CDR3 that has already undergone positive and negative selection, the methodology provides a platform to ask another interesting question. I hypothesize that for a given αβ TCR, if I remove the CDR3 regions and replace them

with short glycine linkers, I might be able to measure the interaction of this TCR with different MHC molecules presenting a shaved peptide with only alanines as the upwardly pointing peptide residues. This measurement would be extremely useful because it will directly measure the germline encoded part of the TCR-pMHC interaction without much input from the CDR3s or the peptide. Should these measurements be quantified, I then could perform alanine scans in the germline encoded CDR1 and CDR2 regions of the TCR as well as the upwardly pointing residues on the MHC helices and clearly characterize all the rules governing the TCR's specificity for MHC once and for all. If, on the other hand, these interactions are not measurable or just as strong as the baseline for interaction with a protein such as CD155, identified to interact with specific TCRs (68), then perhaps there are no germline encoded interactions that bias TCRs towards MHC and all current evidence is some sort of artifact. Either way, an experiment along these lines that is able to divorce the CDR3/peptide interaction from the rest of the TCR-pMHC interaction seems like an obvious choice for the next direction to take these studies.

Concluding Remarks

The evolutionary hypothesis for T cell antigen receptor-peptide major histocompatibility complex (TCR-pMHC) interaction posits the existence of germline-encoded rules by which the TCR is biased towards recognition of the MHC. Understanding these rules is important for our knowledge of how to manipulate this important interaction at the center of adaptive immunity. In this study, I highlight the flexibility of thymic selection as well as existence of these rules by generating knock-in mutant MHC mice and extensively studying their TCR repertoires. Identifying

novel TCR subfamilies most evolutionarily conserved to recognize specific areas of the MHC is the first step in advancing our knowledge of this central interaction.

Evolutionarily, it is important to consider why these solvent exposed amino acids on the α helices of MHC have less polymorphism and some evolutionary conservation. Knowing their importance for interaction with TCRs, one might have thought that the mutant mice would be unable to select such a diverse repertoire, but that was not observed to be the case. Clearly there is a bias in the TCR elements used, but the overall repertoire is far from heavily constrained. The answer as to why these amino acids occur the way they do is probably more complicated and could involve specific responses to specific pathogens as well maintaining homeostasis with the environment. As this is asking an evolutionary question but only in a one generation experiment, it is possible that over time and with the proper selective pressure, the answer would become more obvious.

REFERENCES

1. Little CC & Tyzzer EE (1916) Further experimental studies on the inheritance of susceptibility to a Transplantable tumor, Carcinoma (J. W. A.) of the Japanese waltzing Mouse. *The Journal of medical research* 33(3):393-453.

2. Gorer PA (1936) The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. *Br. J. Exp. PathoL* (17):42-50.

3. Klein J (1975) *Biology of the Mouse Histocompatibility-2 Complex* (Springer-Verlag, New York, NY).

4. Gotze D (1978) The major histocompatibility complex of the rat,RT1 : I. serological characterization of the MNR haplotype (RT1 (m)) in regard to the cross-reacting haplotypesRT1 (a),RT1 (c), andRT1 (b). *Immunogenetics* 7(1):491-506.

5. Owen RD (1945) Immunogenetic Consequences of Vascular Anastomoses between Bovine Twins. *Science* 102(2651):400-401.

6. Billingham RE, Brent L, & Medawar PB (1953) Actively acquired tolerance of foreign cells. *Nature* 172(4379):603-606.

7. McDevitt HO (2000) Discovering the role of the major histocompatibility complex in the immune response. *Annual review of immunology* 18:1-17.

8. Benacerraf B (1991) When all is said and done. *Annual review of immunology* 9:1-26.

9. Klein J, Figueroa F, & Nagy ZA (1983) Genetics of the major histocompatibility complex: the final act. *Annual review of immunology* 1:119-142.

10. Shreffler DC & Owen RD (1963) A Serologically Detected Variant in Mouse Serum: Inheritance and Association with the Histocompatibility-2 Locus. *Genetics* 48(1):9-25.

11. Curman B*, et al.* (1975) H-2 linked Ss protein is C4 component of complement. *Nature* 258(5532):243-245.

12. Claman HN, Chaperon EA, & Triplett RF (1966) Thymus-marrow cell combinations. Synergism in antibody production. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* 122(4):1167-1171.

13. Mitchison NA (1971) The carrier effect in the secondary response to hapten-protein conjugates. V. Use of antilymphocyte serum to deplete animals of helper cells. *European journal of immunology* 1(2):68-75.

14. Senyk G, Nitecki D, & Goodman JW (1971) Immunogenicity of glucagon: determinants responsible for antibody binding and lymphocyte stimulation. *Science* 171(3969):407-408.

15. Hoffmann M & Kappler JW (1973) Regulation of the immune response. II. Qualitative and quantitative differences between thymus- and bone marrow-derived lymphocytes in the recognition of antigen. *The Journal of experimental medicine* 137(3):721-739.

16. Kindred B & Shreffler DC (1972) H-2 dependence of co-operation between T and B cells in vivo. *Journal of immunology* 109(5):940-943.

17. Katz DH, Hamaoka T, Dorf ME, & Benacerraf B (1973) Cell interactions between histoincompatible T and B lymphocytes. The H-2 gene complex determines successful physiologic lymphocyte interactions. *Proceedings of the National Academy of Sciences of the United States of America* 70(9):2624-2628.

18. Rosenthal AS & Shevach EM (1973) Function of macrophages in antigen recognition by guinea pig T lymphocytes. I. Requirement for histocompatible macrophages and lymphocytes. *The Journal of experimental medicine* 138(5):1194-1212.

19. Zinkernagel RM & Doherty PC (1974) Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248(5450):701-702.

20. Katz DH, Graves M, Dorf ME, Dimuzio H, & Benacerraf B (1975) Cell interactions between histoincompatible T and B lymphocytes. VII. Cooperative responses between lymphocytes are controlled by genes in the I region of the H-2 complex. *The Journal of experimental medicine* 141(1):263-268.

21. Kappler JW, Skidmore B, White J, & Marrack P (1981) Antigen-inducible, H-2-restricted, interleukin-2-producing T cell hybridomas. Lack of independent antigen and H-2 recognition. *The Journal of experimental medicine* 153(5):1198-1214.

22. Haskins K*, et al.* (1983) The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. *The Journal of experimental medicine* 157(4):1149-1169.

23. Allison JP, McIntyre BW, & Bloch D (1982) Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *Journal of immunology* 129(5):2293-2300.

24. Meuer SC*, et al.* (1983) Clonotypic structures involved in antigen-specific human T cell function. Relationship to the T3 molecular complex. *The Journal of experimental medicine* 157(2):705-719.

25. Hedrick SM, Cohen DI, Nielsen EA, & Davis MM (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308(5955):149-153.

26. Yanagi Y*, et al.* (1984) A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308(5955):145-149.

27. Ellner JJ, Lipsky PE, & Rosenthal AS (1977) Antigen handling by guinea pig macrophages: further evidence for the sequestration of antigen relevant for activation of primed T lymphocytes. *Journal of immunology* 118(6):2053-2057.

28. Ziegler K & Unanue ER (1981) Identification of a macrophage antigen-processing event required for I-region-restricted antigen presentation to T lymphocytes. *Journal of immunology* 127(5):1869-1875.

29. Shimonkevitz R, Kappler J, Marrack P, & Grey H (1983) Antigen recognition by H-2-restricted T cells. I. Cell-free antigen processing. *The Journal of experimental medicine* 158(2):303-316.

30. Babbitt BP, Allen PM, Matsueda G, Haber E, & Unanue ER (1985) Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 317(6035):359-361.

31. Buus S, Sette A, Colon SM, Jenis DM, & Grey HM (1986) Isolation and characterization of antigen-Ia complexes involved in T cell recognition. *Cell* 47(6):1071-1077.

32. Bjorkman PJ*, et al.* (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329(6139):506-512.

33. Bjorkman PJ*, et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329(6139):512-518.

34. Brown JH*, et al.* (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364(6432):33-39.

35. Paul WE (2008) *Fundamental immunology* (Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia) 6th Ed pp xviii, 1603 p., 1616 p. of plates.

36. Townsend AR*, et al.* (1986) The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides. *Cell* 44(6):959-968.

37. Marrack P & Kappler J (1987) The T cell receptor. *Science* 238(4830):1073-1079.

38. Davis MM & Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. *Nature* 334(6181):395-402.

39. Garcia KC*, et al.* (1996) An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science* 274(5285):209-219.

40. Jorgensen JL, Esser U, Fazekas de St Groth B, Reay PA, & Davis MM (1992) Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature* 355(6357):224-230.

41. Sant'Angelo DB*, et al.* (1996) The specificity and orientation of a TCR to its peptide-MHC class II ligands. *Immunity* 4(4):367-376.

42. Rudolph MG, Stanfield RL, & Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annual review of immunology* 24:419-466.

43. Marrack P, Scott-Browne JP, Dai S, Gapin L, & Kappler JW (2008) Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annual review of immunology* 26:171-203.

44. Jerne NK (1971) The somatic generation of immune recognition. *European journal of immunology* 1(1):1-9.

45. Fink PJ & Bevan MJ (1978) H-2 antigens of the thymus determine lymphocyte specificity. *The Journal of experimental medicine* 148(3):766-775.

46. Wilson IA & Stanfield RL (2005) MHC restriction: slip-sliding away. *Nature immunology* 6(5):434-435.

47. Sim BC*, et al.* (1998) Thymic skewing of the CD4/CD8 ratio maps with the T-cell receptor alpha-chain locus. *Current biology : CB* 8(12):701-704.

48. Sim BC, Zerva L, Greene MI, & Gascoigne NR (1996) Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science* 273(5277):963-966.

49. Bill J, Appel VB, & Palmer E (1988) An analysis of T-cell receptor variable region gene expression in major histocompatibility complex disparate mice. *Proceedings of the National Academy of Sciences of the United States of America* 85(23):9184-9188.

50. Sullivan BA & Kronenberg M (2007) TCR-mediated recognition of glycolipid CD1 complexes. *Current topics in microbiology and immunology* 314:165-193.

51. Zerrahn J, Held W, & Raulet DH (1997) The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* 88(5):627-636.

52. Merkenschlager M*, et al.* (1997) How many thymocytes audition for selection? *The Journal of experimental medicine* 186(7):1149-1158.

53. McDonald BD, Bunker JJ, Erickson SA, Oh-Hora M, & Bendelac A (2015) Crossreactive alphabeta T Cell Receptors Are the Predominant Targets of Thymocyte Negative Selection. *Immunity* 43(5):859-869.

54. Felix NJ & Allen PM (2007) Specificity of T-cell alloreactivity. *Nature reviews. Immunology* 7(12):942-953.

55. Colf LA*, et al.* (2007) How a single T cell receptor recognizes both self and foreign MHC. *Cell* 129(1):135-146.

56. Clark SP, Arden B, Kabelitz D, & Mak TW (1995) Comparison of human and mouse T-cell receptor variable gene segment subfamilies. *Immunogenetics* 42(6):531-540.

57. Garcia KC, Teyton L, & Wilson IA (1999) Structural basis of T cell recognition. *Annual review of immunology* 17:369-397.

58. Beringer DX*, et al.* (2015) T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nature immunology* 16(11):1153-1161.

59. Adams JJ*, et al.* (2011) T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* 35(5):681-693.

60. Kappler JW, Roehm N, & Marrack P (1987) T cell tolerance by clonal elimination in the thymus. *Cell* 49(2):273-280.

61. Kisielow P, Bluthmann H, Staerz UD, Steinmetz M, & von Boehmer H (1988) Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature CD4+8+ thymocytes. *Nature* 333(6175):742-746.

62. Ignatowicz L, Kappler J, & Marrack P (1996) The repertoire of T cells shaped by a single MHC/peptide ligand. *Cell* 84(4):521-529.

63. Dai S*, et al.* (2008) Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity* 28(3):324-334.

64. Feng D, Bond CJ, Ely LK, Maynard J, & Garcia KC (2007) Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nature immunology* 8(9):975-983.

65. Maynard J*, et al.* (2005) Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity* 22(1):81-92.

66. Scott-Browne JP, White J, Kappler JW, Gapin L, & Marrack P (2009) Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature* 458(7241):1043-1046.

67. Van Laethem F*, et al.* (2007) Deletion of CD4 and CD8 coreceptors permits generation of alphabetaT cells that recognize antigens independently of the MHC. *Immunity* 27(5):735-750.

68. Tikhonova AN*, et al.* (2012) alphabeta T cell receptors that do not undergo major histocompatibility complex-specific thymic selection possess antibody-like recognition specificities. *Immunity* 36(1):79-91.

69. Van Laethem F*, et al.* (2013) Lck availability during thymic selection determines the recognition specificity of the T cell repertoire. *Cell* 154(6):1326-1341.

70. Holland SJ*, et al.* (2012) The T-cell receptor is not hardwired to engage MHC ligands. *Proceedings of the National Academy of Sciences of the United States of America* 109(45):E3111-3118.

71. Burrows SR*, et al.* (2010) Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proceedings of the National Academy of Sciences of the United States of America* 107(23):10608-10613.

72. Stadinski BD*, et al.* (2011) A role for differential variable gene pairing in creating T cell receptors specific for unique major histocompatibility ligands. *Immunity* 35(5):694-704.

73. Garcia KC*, et al.* (2012) A closer look at TCR germline recognition. *Immunity* 36(6):887-888.

74. Anonymous (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 401(6756):921-923.

75. Janeway CA, Jr. (1989) Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor symposia on quantitative biology* 54 Pt 1:1-13.

76. Kalbe M*, et al.* (2009) Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings. Biological sciences / The Royal Society* 276(1658):925-934.

77. Rossjohn J*, et al.* (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annual review of immunology* 33:169-200.

78. Bartl S & Weissman IL (1994) Isolation and characterization of major histocompatibility complex class IIB genes from the nurse shark. *Proceedings of the National Academy of Sciences of the United States of America* 91(1):262-266.

79. Kaufman JF, Auffray C, Korman AJ, Shackelford DA, & Strominger J (1984) The class II molecules of the human and murine major histocompatibility complex. *Cell* 36(1):1-13.

80. Huseby ES*, et al.* (2005) How the T cell repertoire becomes peptide and MHC specific. *Cell* 122(2):247-260.

81. Scott-Browne JP*, et al.* (2011) Evolutionarily conserved features contribute to alphabeta T cell receptor specificity. *Immunity* 35(4):526-535.

82. Rubtsova K*, et al.* (2009) Many different Vbeta CDR3s can reveal the inherent MHC reactivity of germline-encoded TCR V regions. *Proceedings of the National Academy of Sciences of the United States of America* 106(19):7951-7956.

83. Swift S, Lorens J, Achacoso P, & Nolan GP (2001) Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. *Current protocols in immunology / edited by John E. Coligan ... [et al.]* Chapter 10:Unit 10 17C.

84. Glimcher LH, McKean DJ, Choi E, & Seidman JG (1985) Complex regulation of class II gene expression: analysis with class II mutant cell lines. *Journal of immunology* 135(5):3542-3550.

85. White J*, et al.* (1989) Two better cell lines for making hybridomas expressing specific T cell receptors. *Journal of immunology* 143(6):1822-1825.

86. Nagy A, Gertsenstein M, Vintersten K, & Behringer R (2006) Collecting two-cell- to compacted morula-stage embryos. *CSH protocols* 2006(1).

87. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):550.

88. Anonymous (2005) Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

89. Kostic AD*, et al.* (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome research* 22(2):292-298.

90. Decock A*, et al.* (2012) Genome-wide promoter methylation analysis in neuroblastoma identifies prognostic methylation biomarkers. *Genome biology* 13(10):R95.

91. Taiwo O*, et al.* (2013) DNA methylation analysis of murine hematopoietic side population cells during aging. *Epigenetics* 8(10):1114-1122.

92. Stark R & Brown G (2013) DiffBind: differential binding analysis of ChIP-Seq peak data.

93. Ross-Innes CS*, et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481(7381):389-393.

94. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289-300.

95. Weiner J (2015) pca3d: Three Dimensional PCA Plots. in *R package version 0.8* (http://CRAN.R-project.org/package=pca3d).

96. Cohn LE*, et al.* (1986) Identification of functional regions on the I-Ab molecule by site-directed mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* 83(3):747-751.

97. Doebele RC, Busch R, Scott HM, Pashine A, & Mellins ED (2000) Determination of the HLA-DM interaction site on HLA-DR molecules. *Immunity* 13(4):517-527.

98. Lippolis JD*, et al.* (2002) Analysis of MHC class II antigen processing by quantitation of peptides that constitute nested sets. *Journal of immunology* 169(9):5089-5097.

99. Liu X*, et al.* (2002) Alternate interactions define the binding of peptides to the MHC molecule IA(b). *Proceedings of the National Academy of Sciences of the United States of America* 99(13):8820-8825.

100. Dongre AR*, et al.* (2001) In vivo MHC class II presentation of cytosolic proteins revealed by rapid automated tandem mass spectrometry and functional analyses. *European journal of immunology* 31(5):1485-1494.

101. Kato Z*, et al.* (2008) Positioning of autoimmune TCR-Ob.2F3 and TCR-Ob.3D1 on the MBP85-99/HLA-DR2 complex. *Proceedings of the National Academy of Sciences of the United States of America* 105(40):15523-15528.

102. Kouskoff V, Fehling HJ, Lemeur M, Benoist C, & Mathis D (1993) A vector driving the expression of foreign cDNAs in the MHC class II-positive cells of transgenic mice. *Journal of immunological methods* 166(2):287-291.

103. Kim HJ, Guo D, & Sant'Angelo DB (2005) Coevolution of TCR-MHC interactions: conserved MHC tertiary structure is not sufficient for interactions with the TCR. *Proceedings of the National Academy of Sciences of the United States of America* 102(20):7263-7267.

104. Carbery ID*, et al.* (2010) Targeted genome modification in mice using zinc-finger nucleases. *Genetics* 186(2):451-459.

105. Remy S*, et al.* (2010) Zinc-finger nucleases: a powerful tool for genetic engineering of animals. *Transgenic research* 19(3):363-371.

106. Chen F*, et al.* (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature methods* 8(9):753-755.

107. Urnov FD*, et al.* (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435(7042):646-651.

108. Boch J (2011) TALEs of genome targeting. *Nature biotechnology* 29(2):135-136.

109. Pennisi E (2013) The CRISPR craze. *Science* 341(6148):833-836.

110. Pelletier S, Gingras S, & Green DR (2015) Mouse genome engineering via CRISPR-Cas9 for study of immune function. *Immunity* 42(1):18-27.

111. Zhang L*, et al.* (2015) Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PloS one* 10(3):e0120396.

112. Lefranc MP*, et al.* (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic acids research* 37(Database issue):D1006-1012.

113. Fisher RA, Corbet AS, & Williams CB (1943) The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology* 12(1):42-58.

114. Jaccard P (1901) Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37:241-272.

115. Chao A, Chazdon RL, Colwell RK, & Shen T-J (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* 8(2):148-159.

116. Qi Q*, et al.* (2014) Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences of the United States of America* 111(36):13139-13144.

117. Roth ME, Holman PO, & Kranz DM (1991) Nonrandom use of J alpha gene segments. Influence of V alpha and J alpha gene location. *Journal of immunology* 147(3):1075-1081.

118. Huang C & Kanagawa O (2001) Ordered and coordinated rearrangement of the TCR alpha locus: role of secondary rearrangement in thymic selection. *Journal of immunology* 166(4):2597-2601.

119. Nemazee D & Hogquist KA (2003) Antigen receptor selection by editing or downregulation of V(D)J recombination. *Current opinion in immunology* 15(2):182-189.

120. Gavin MA & Bevan MJ (1995) Increased peptide promiscuity provides a rationale for the lack of N regions in the neonatal T cell repertoire. *Immunity* 3(6):793-800.

121. Deng L, Langley RJ, Wang Q, Topalian SL, & Mariuzza RA (2012) Structural insights into the editing of germ-line-encoded interactions between T-cell receptor and MHC class II by Valpha CDR3. *Proceedings of the National Academy of Sciences of the United States of America* 109(37):14960-14965.

122. Cole DK*, et al.* (2014) T-cell receptor (TCR)-peptide specificity overrides affinity-enhancing TCR-major histocompatibility complex interactions. *The Journal of biological chemistry* 289(2):628-638.

123. Yin L, Scott-Browne J, Kappler JW, Gapin L, & Marrack P (2012) T cells and their eons-old obsession with MHC. *Immunological reviews* 250(1):49-60.

124. Kosmrlj A, Jha AK, Huseby ES, Kardar M, & Chakraborty AK (2008) How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proceedings of the National Academy of Sciences of the United States of America* 105(43):16671-16676.

125. Adams JJ*, et al.* (2016) Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nature immunology* 17(1):87-94.

126. Reantragoon R*, et al.* (2012) Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor. *The Journal of experimental medicine* 209(4):761-774.

127. Gadola SD*, et al.* (2006) Structure and binding kinetics of three different human CD1d-alpha-galactosylceramide-specific T cell receptors. *The Journal of experimental medicine* 203(3):699-710.

128. Wilson IA & Bjorkman PJ (1998) Unusual MHC-like molecules: CD1, Fc receptor, the hemochromatosis gene product, and viral homologs. *Current opinion in immunology* 10(1):67-73.

129. Hahn M, Nicholson MJ, Pyrdol J, & Wucherpfennig KW (2005) Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nature immunology* 6(5):490-496.

130. Hanada K, Wang QJ, Inozume T, & Yang JC (2011) Molecular identification of an MHC-independent ligand recognized by a human {alpha}/{beta} T-cell receptor. *Blood* 117(18):4816-4825.

131. Mallis RJ*, et al.* (2015) Pre-TCR ligand binding impacts thymocyte development before alphabetaTCR expression. *Proceedings of the National Academy of Sciences of the United States of America* 112(27):8373-8378.