Published in final edited form as:
Cell. 2014 May 22; 157(5): 1073–1087. doi:10.1016/j.cell.2014.03.047.

Deconstructing the peptide-MHC specificity of T cell recognition

Michael E. Birnbaum[1,3], **Juan L. Mendoza**[1], **Dhruv K. Sethi**[5], **Shen Dong**[1], **Jacob Glanville**[2,3], **Jessica Dobbins**[5,6], **Engin Özkan**[1,4], **Mark M. Davis**[2,3,4], **Kai W. Wucherpfennig**[5,6], and **K. Christopher Garcia**[1,3,4]

[1]Departments of Molecular and Cellular Physiology and Structural Biology, Stanford University School of Medicine, Stanford University, Stanford, CA 94305

[2]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford University, Stanford, CA 94305

[3]Program in Immunology, Stanford University School of Medicine, Stanford University, Stanford, CA 94305

[4]The Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305

[5]Department of Cancer Immunology & AIDS, Dana-Farber Cancer Institute, Boston, MA 02115

[6]Program in Immunology, Harvard Medical School, Boston, MA 02115

Summary

In order to survey a universe of MHC-presented peptide antigens whose numbers greatly exceed the diversity of the T cell repertoire, T cell receptors (TCRs) are thought to be cross-reactive. However, the nature and extent of TCR cross-reactivity has not been conclusively measured experimentally. We developed a system to identify MHC-presented peptide ligands by combining TCR selection of highly diverse yeast-displayed peptide-MHC libraries with deep sequencing. While we identified hundreds of peptides reactive with each of five different mouse and human TCRs, the selected peptides possessed TCR recognition motifs that bore a close resemblance to their known antigens. This structural conservation of the TCR interaction surface allowed us to exploit deep sequencing information to computationally identify activating microbial and self-ligands for human autoimmune TCRs. The mechanistic basis of TCR cross-reactivity described

Address correspondence to: K. Christopher Garcia, kcgarcia@stanford.edu Tel# 650-498-7332.

Accession numbers The coordinates and structure factors for the reported crystal structures are deposited in the Protein Data Bank (PDB) under PDB IDs 4P2O, 4P2Q, and 4P2R. Deep sequencing data can be accessed via the Sequence Read Archive (SRA) under project code PRJNA240193.

Contributions M.E.B. and K.C.G. conceived of the project and wrote the manuscript. M.E.B. and S.D. designed and performed yeast selections. M.E.B., J.L.M., and J.G. conducted deep sequencing analysis. J.L.M. authored the peptide search algorithm. M.E.B. and D.K.S. expressed recombinant proteins. M.E.B., D.K.S., and J.D. validated peptide binding and activity. M.E.B. and E.O. performed crystallization experiments of pMHC-TCR complexes. M.M.D., K.W.W., and K.C.G. supervised the research. All authors edited the manuscript.

here enables effective surveillance of diverse self and foreign antigens, but without necessitating degenerate recognition of non-homologous peptides.

Introduction

T cells are central to many aspects of adaptive immunity. Each mature T cell expresses a unique αβ T cell receptor (TCR) that has been selected for its ability to bind to peptides presented by major histocompatibility complex (MHC) molecules. Unlike antibodies, TCRs generally have low affinity for ligands (KD ~ 1-100 μM), which has been speculated to facilitate rapid scanning of peptide-MHC (pMHC) (Matsui et al., 1991; Rudolph et al., 2006; Wu et al., 2002). Structural studies of TCR-pMHC complexes have revealed a binding orientation where, generally, the TCR CDR1 and CDR2 loops make the majority of contacts with the tops of the MHC helices while the CDR3 loops, which are conformationally malleable, primarily engage the peptide presented in the MHC groove (Davis and Bjorkman, 1988; Garcia and Adams, 2005; Rudolph et al., 2006).

The low affinity and fast kinetics of TCR-pMHC binding, combined with conformational plasticity in the CDR3 loops, would seem to facilitate cross-reactivity with structurally distinct peptides presented by MHC (Mazza et al., 2007; Reiser et al., 2003; Yin and Mariuzza, 2009). Indeed, given that the calculated diversity of potential peptide antigens is much larger than TCR repertoire diversity, TCR cross-reactivity appears to be a biological imperative (Mason, 1998; Wooldridge et al., 2012). Cross-reactive TCRs have been implicated in both pathogenic and protective roles for a number of diseases (Benoist and Mathis, 2001; De la Herran-Arita et al., 2013; Shann et al., 2010; Welsh et al., 2010; Wucherpfennig and Strominger, 1995).

Nevertheless, the true extent of TCR cross-reactivity, and its role in T cell immunity, remains a speculative issue, largely due to the absence of quantitative experimental approaches that could definitively address this question (Mason, 1998; Morris and Allen, 2012; Shih and Allen, 2004; Wilson et al., 2004; Wucherpfennig et al., 2007). While many examples exist of TCRs recognizing substituted or homologous peptides related to the antigen (Krogsgaard et al., 2003), such as altered peptide ligands (Kersh and Allen, 1996), most of these peptides retain similarities to the wild-type peptides and are recognized in a highly similar fashion. Only a handful of defined examples exist of a single TCR recognizing non-homologous sequences (Adams et al., 2011; Basu et al., 2000; Colf et al., 2007; Ebert et al., 2009; Evavold et al., 1995; Lo et al., 2009; Macdonald et al., 2009; Nanda et al., 1995; Reiser et al., 2003; Zhao et al., 1999).

One approach that has been used to estimate cross-reactivity utilizes pooled, chemically synthesized peptide libraries (Hemmer et al., 1998b; Wilson et al., 2004; Wooldridge et al., 2012). Using calculations based upon this technique, it has been extrapolated that ~10^6 different peptides in mixtures containing ~10^{12} different peptides were agonists (Wilson et al., 2004; Wooldridge et al., 2012). Synthetic peptide libraries have been used to isolate diverse peptide sequences (Hemmer et al., 1998a), including microbial and self ligands for TCRs of interest (Hemmer et al., 1997). However, most studies find only close homologues to known peptides (Krogsgaard et al., 2003; Maynard et al., 2005; Wilson et al., 1999;

Wilson et al., 2004). Furthermore, these libraries were assayed based solely on bulk stimulatory ability with femtomolar concentrations of any given peptide and no knowledge of peptide loading in the MHC or pMHC binding to the TCR. A more accurate estimate of cross-reactivity requires the isolation of individual sequences from a library of MHC-presented peptides based upon *binding* to a TCR.

Recently, we and others have created libraries of peptides linked to MHC via yeast and baculovirus display as a method to discover TCR ligands through affinity-based selections that rely on a physical interaction between the pMHC and the TCR (Adams et al., 2011; Birnbaum et al., 2012; Crawford et al., 2004; Crawford et al., 2006; Macdonald et al., 2009; Wang et al., 2005). However, these methods have so far not been used to address the broader question of TCR cross-reactivity, since the requirement of manually validating and sequencing individual library 'hits' has restricted the approach to discovering small numbers of peptides.

Here, we combined affinity-based selections of peptide-MHC yeast libraries and deep sequencing to discover hundreds of unique peptide sequences recognized by multiple murine and human TCRs. Strikingly, all peptide sequences bear TCR epitopes with close similarity to their previously known agonist antigens. With an understanding of this property, we created a computational algorithm to predict naturally occurring TCR ligands using data from our deep sequencing results. We tested a diverse set of the putative TCR-reactive peptides and found 94% are able to elicit a T cell response. In general, TCR cross-reactivity does not appear to be characterized by broad degeneracy, but rather is largely constrained to a small number of TCR contact residue 'hotspots' on a peptide, while tolerating extensive diversity at other positions. This more granular understanding of the properties of TCR cross-reactivity has broad implications for ligand identification, vaccine design, and immunotherapy.

Results

Development and selection of a murine MHC platform for yeast display

We developed a system for the rapid and sensitive detection of TCR-binding peptides presented by the murine class II MHC I-E^k. This represents an advance over previous reports of class II pMHC molecules displayed on the surface of yeast that either did not show or were not tested for the ability to bind soluble TCR (Birnbaum et al., 2012; Boder et al., 2005; Esteban and Zhao, 2004; Jiang and Boder, 2010; Starwalt et al., 2003; Wen et al., 2008; Wen et al., 2011). We designed our construct as a 'mini' single-chain MHC Aga2 fusion, with the truncated peptide-binding β1α1 domains fused and the wild-type peptide MCC fused to the N-terminus of the MHC β1 domain via Gly-Ser linkers (Figure 1A) (Adams et al., 2011). The initial construct was correctly routed to the yeast surface but did not have the ability to bind to TCR, indicating the pMHC was not correctly folded (Figure 1B).

In order to rescue folding of the pMHC, we subjected the mini I-E^k to error-prone mutagenesis combined with introduction of solubility-enhancing mutations. We selected this mutagenized mini scaffold for binding to the 2B4 TCR, which recognizes MCC-I-E^k with

moderate affinity and slow kinetics (Newell et al., 2011). Our final construct contained solubilizing mutations in what was previously the α1β1-α2β2 domain interface and one mutation between the MHC helix and the beta sheets (Figure 1B). None of the mutated MHC residues contacted either the peptide or the TCR. The evolved construct retained specific binding to several MCC-I-E^k recognizing TCRs and showed comparable affinity to the wild-type pMHC (Figure 1B, S1).

We then created a peptide library tethered to the MHC construct. Based upon the recently solved 2B4-MCC-I-E^k structure (Newell et al., 2011), we mutagenized the peptide from P(-2) to P10 (Figure 1C). Limited diversity was introduced at the two most distal residues and the primary MHC-binding anchor residues at P1 and P9 to maximize the number of peptides capable of being correctly displayed by the MHC (Figure 1C).

Our first attempts at screening involved ‘manual curation’ of selections conducted with multivalent TCR. The library showed enrichment after three rounds of selection using highly avid TCR-coated streptavidin beads followed by a higher stringency ‘polishing’ round of selection using TCR tetramers. The three peptides recovered via sequencing of 12 individual, hand picked clones after selection were related to the WT MCC peptide – the P2, P5, and P8 TCR contacts were all conserved, while P3 showed a conservative Tyr to Phe mutation (Figure 1D). We surmised that these enriched WT-like sequences present in the later rounds dominated the selections, preventing alternative, potentially non-homologous sequences from being recovered. For this reason, we turned to deep sequencing at each step of the selection process to recover all enriched peptides.

Deep sequencing of selections for TCR-binding peptides

Analysis of the pooled yeast library DNA after each successive round of selection with 2B4 via deep sequencing showed enrichment from an essentially random distribution of amino acids to a highly WT-like TCR recognition motif (Figures 2A, S2A). After the third round, there were non-homologous amino acids at P5 and P8 selected above background (Met and Ser for P5, Ile and Leu for P8) that were outcompeted by the WT-like motif by the final round of selection. Overall, the number of unique peptides observed via deep sequencing progressed from 132,000 unique in-frame peptides observed in the sequenced portion of pre-selection library to only 207 unique peptides after the 3rd round of selection (Figures 2B, 2C, S2A, S2B). By the final round of selection, the library was dominated by a handful of sequences, matching the result obtained by manual curation (Figures 1D, 2B, 2C).

We repeated the selections with two other TCRs reactive to MCC-I-E^k: 226 and 5cc7. We analyzed enrichment for each TCR after the third round of selection, where there is enrichment for a binding motif but before complete convergence to a small number of sequences (Figures 2A, 3A, S2B, S3A). While all three TCRs retain a WT-like TCR recognition motif (indicated by the outlined boxes in the heatmaps), each TCR also shows some variation in positional preferences (Figure 3A). For example, where 2B4 can recognize P5 Met and Ser, 5cc7 can accommodate P5 Leu, Val, and Arg. The P3 TCR contact position showed the least variance across all three TCRs, with either Phe or Tyr being required for 2B4 and 5cc7, and Phe, Tyr, or Trp being required for 226 (Figure 3A). 226, as previously reported, showed a greater degree of cross-reactivity, able to recognize 897 unique peptide

sequences. The larger number of peptides recognized was largely a function of a higher tolerance for substitutions on TCR-neutral and MHC-contacting residues, such as at positions P(-1) and P4 (Figure 3A, S3A) (Ehrich et al., 1993; Newell et al., 2011).

The large collection of peptides recovered via deep sequencing enabled us to apply covariation analysis to discover intra-peptide structure-activity relationships that were not previously accessible with traditional single residue substitution analysis (Figure 3B) (Ehrich et al., 1993; Newell et al., 2011; Reay et al., 1994; Wilson et al., 1999). By using co-variation analysis of the central P5 residue and the C-terminal P8 residue, a pattern emerged: the native, MCC-like 'up-facing' TCR-contact motifs for each TCR (P5 Lys, P8 Ser/Thr) were strongly correlated, while the altered residues (P5 Ser/P8Leu for 2B4, P5 Leu or Arg/P8 Phe for 5cc7) independently segregated (Figure 3B). These results highlight a degree of cooperativity in the composition of residues comprising a 'TCR epitope' that is clearly revealed with deep sequencing. Furthermore, such intra-peptide residue coupling reveals how cross-reactivity can occur through mutually compensatory substitutions to the parent peptide.

While the selected ligands for all three TCRs possessed shared features, each TCR also selected for a subset of sequences that were not selected by the other two. We applied distance clustering to the peptides selected by all three TCRs to determine if all selected sequences were part of the larger MCC-like peptide family or were distinct families (Figure 3C). We found that while sequences recognized by individual TCRs clustered more closely to each other, essentially all of the selected sequences formed one large cluster of peptides no more than three amino acids different than at least one other peptide in the cluster (Figure 3C, S3B). Therefore, the selected peptides for all three TCRs are related via a common specificity domain, and importantly, to the parent MCC ligand. Even though we conducted unbiased selections of random libraries, the only ligands that were recovered were remarkably similar to the WT ligand at the TCR interface.

Functional characterization of I-E^k library hits

We synthesized 44 of the library peptides selected for binding to the TCRs and examined their ability to stimulate T cell blasts from 2B4 and 5cc7 transgenic mice as assayed by CD69 upregulation and IL-2 production. The majority of the peptides predicted to bind 2B4 (19/19) and 5cc7 (17/21) expressing T cells induced CD69 upregulation (Figures 4A, 4B, S4A-D). The peptides had a wide range of potencies, including ~50-fold more potent than the wild-type peptide MCC (colored red). When we compared the presence of the MCC-like TCR recognition epitope with TCR signaling, we found that in general, sequences that shared the MCC-like epitope at all three major TCR contacts (colored blue) were more potent in inducing signaling than those peptides that were more distantly related (colored black) (Figures 4A, 4B). We also tested the peptides selected for binding to one TCR for their ability to cross-react with the other MCC-reactive T cells. Surprisingly, a large proportion of these peptides potently activated TCR signaling (Figures 4A, 4B, S4A-D). In general, the sequences that showed the most robust activation were again the ones that most closely shared the MCC TCR binding epitope.

We additionally chose nine peptides from our initial set of 44 and exchanged them into soluble I-E^k MHC for TCR affinity measurements via surface plasmon resonance (SPR). For 2B4 and 5cc7, TCR bound the pMHC of interest with affinities ranging from KD of ~1 μM (over 10-fold better than MCC) to those with binding only barely detectable at 100 μM TCR (Figure S4E-F). When we compared the activity and affinity of our selected peptides, there is a loose but positive correlation between strength of TCR-pMHC binding and potency of activation (Figure 4C).

The structural basis of TCR recognition of cross-reactive peptides

To determine the molecular basis of the TCRs' ability to recognize the most diverse peptides selected from our I-E^k libraries, we determined the crystal structures of 2B4 in complex with a peptide termed 2A bound to I-E^k, as well as 5cc7 in complex with two peptides bound to IE^k, termed 5c1 and 5c2 (Table S1, Figures 5A and 5B). When these complexes were aligned with previously solved complex structures of TCRs (2B4 and 226) binding to MCC-I-E^k, very little deviation in overall TCR-pMHC complex geometry from the parent complexes was observed (Figures 5A and 5B) (Newell et al., 2011). Since the 5cc7-MCC-I-E^k complex is not solved, 5c1 and 5c2 were compared to 226-MCC-I-E^k, which shares the TCRβ chain with 5cc7 and therefore likely retains a close footprint. The contacts between TCR germline-derived CDR1/2 loops and MHC helices, which make up roughly 50% of the binding interface between TCR and pMHC, were essentially unchanged in the new peptide complexes versus MCC (Figure 5C).

When we examined the chemistry of MCC versus 2A, and MCC versus 5c1 peptide recognition by their respective TCRs, we saw the interaction between the TCRα CDR loops and the N-terminal half of the peptides are essentially invariant (Figures 5A and 5B, lower panels). Each peptide backbone makes a hydrogen bond at the P3 carbonyl with Arg29α in the TCR CDR1α loop (Figure S5A). The contacts of 2B4 CDR3α with P2 and P3 in MCC and 2A are essentially identical (Figure 5A, lower panels). While an exact analogy cannot be made between 5cc7 recognizing 5c1 and 226 recognizing MCC due to sequence differences in their CDR3 loops, 5cc7 and 226 CDR3α loop conformations and peptide contacts are extremely similar (Figure 5B, lower panels).

In contrast, 2B4 and 5cc7 β chain CDR loop interactions with the C terminal half of the peptides show marked changes to accommodate the non-MCC sequences. For 2B4, the CDR3β loop conformation completely rearranges to engage the alternate P5 and P8 residues on the 2A peptide (Figure 5A, lower panels). Gln100β, a residue that makes no contact with the peptide in the 2B4-MCC-I-E^k complex, flips its side chain by 180 degrees to form hydrogen bonds with the peptide backbone carbonyl oxygens at P5 and P6 (Figure 5A, lower panels). The side chains of Trp98β and Ser99β form hydrogen bonds with the P5 Ser hydroxyl moiety (Figure 5A). Asp101β, one of the main contacts with P5 Lys in MCC, forms a hydrogen bond with Ser95β on the other end of the CDR3β loop, significantly altering the overall topology of the loop (Figure S5B).

In the 5cc7-5c1-I-E^k complex, there are fewer hydrogen bonds formed between the peptide and TCR due to the replacement of P5 Lys with Leu in the 5c1 peptide (Figure 5B, lower panels). Asn98β changes its hydrogen bonding network from engaging only the carbonyl of

P6 on the MCC peptide backbone to simultaneously interacting with the carbonyl oxygen of P6 and the amide nitrogen of P8 of the 5c1 peptide (Figure 5B). The second 5cc7-reactive peptide, 5c2, is recognized essentially identically by 5cc7 as 5c1 despite the substitution of P5 to Arg (Figure S5C). The substitution of a bulkier side chain at P8 (Phe instead of Thr) results in a rocking of 5cc7 such that the TCR Cβ FG loop is translated by 15 Å relative to the 226-MCC structure (Figure S5D-E). It is interesting to note that all tested peptides with P8 Phe signal less efficiently than MCC-like peptides, even when affinities are closely matched (Figure S4E-F). These structures raise the question if a minor tilt of the TCR relative to the MHC can have consequences for signaling.

Upon closer inspection, we find that homologies between what appear to be unrelated peptide sequences emerge from sequence clustering and structural analysis. For example, close structural relationships between the interaction modes of the 2B4-reactive peptides MCC and 2A are apparent even though the peptides show little homology at 4/5 TCR contact positions (Figure 5A). We set out to determine if we could identify intermediate sequences that would 'evolutionarily' link these two peptide sequences, given that both reside in the same sequence cluster (Figure 3C). Using our dataset of peptide sequences selected for 2B4 binding, we were able to populate a family of peptides that would incrementally link MCC and 2A, with each peptide differing by only one TCR contact from the peptide before and after it (Figure 5D). Thus, connectivity can be established between MCC and 2A through stepwise single amino acid drifts from their parent sequences.

Collectively, despite differences in peptide sequences, all MCC and library-peptide derived complexes share many common features with regards to docking geometry and interaction chemistry. Changes in up-facing peptide residue sequence (e.g. P5, P8) are accommodated 'locally' in a structurally parsimonious fashion that preserves most of the parent MCC peptide complex features, as opposed to accommodation through large scale repositioning of the CDR loops on the pMHC surface.

Development and selection of a human MHC platform for yeast display

To exploit our technology to find ligands for TCRs relevant to human disease, we also engineered the human MHC HLA-DR15, an allele with genetic linkage to multiple sclerosis (Hafler et al., 2007; Patsopoulos et al., 2013). For yeast surface display, HLA-DR15 was constructed comparably to the murine I-E^k β1α1 'mini' MHC (Figure 6A). We chose to examine two closely-related TCRs, Ob.1A12 and Ob.2F3, that were cloned from a patient with relapsingremitting multiple sclerosis and recognize HLA-DR15 bound to an immunodominant epitope of myelin basic protein (MBP, residues 85-99) (Wucherpfennig et al., 1994b). These two TCRs utilize the same Vα-Jα and Vβ-Jβ gene segments and differ at one position in the CDR3α loop and two positions in CDR3β. Ob.1A12 is sufficient to cause disease in a humanized TCR transgenic mouse model (Harkiolaki et al., 2009; Hausmann et al., 1999; Madsen et al., 1999). A structure of Ob.1A12 complexed with MBP-HLA-DR15 revealed an atypical docking mode, with the TCR shifted towards the N-terminus of the peptide and primarily interacting with a P2-His/P3-Phe TCR contact motif (Figure 6A) (Hahn et al., 2005; Wucherpfennig et al., 1994a).

Since the initial wild-type MBP-HLA-DR15 yeast display construct was not stained by Ob.1A12 TCR tetramers, we subjected the construct to error prone mutagenesis and selected for binding with Ob.1A12. Our final construct combined the most heavily selected mutation (Pro11Ser on HLA-DR15β) with two solubility-enhancing mutations on the bottom of the MHC platform (Figure S6A). The final construct stained robustly with Ob.1A12 and Ob.2F3 TCR tetramers (Figure S6B).

We designed a peptide library within the HLA-DR15 mini MHC scaffold to find novel Ob.1A12- and Ob.2F3-reactive peptides (Figure 6A). Since Ob.1A12 binds its cognate pMHC shifted towards the N terminus of the peptide, we extended the library, randomizing from P(-4) to P10 (Hahn et al., 2005). The P1 and P4 peptide anchors for HLA-DR15 were afforded limited diversity. When we selected with Ob.1A12 and Ob.2F3 TCRs, we observed a strong convergence to a wild-type MBP-like TCR recognition motif for the primary TCR contacts (P2 His, P3 Phe, and P5 Lys) (Figure 6B, S6C, S6D).

Given the dominance of the ‘HF’ motif in the selection results, we sought to determine if alternative cross-reactive TCR epitopes would emerge if the motif were suppressed. We made a library that allowed every amino acid except for His at P2, Phe at P3, and Lys at P5 (Figure 6C). After selection, the TCR-binding clones still converged to a central HF motif by register shifting towards the C-terminus of the peptide by one amino acid, allowing the previous P4 Phe anchor to be repurposed as the P3 TCR contact, and the P3 position of the library to become the new P2 His TCR contact (Figure 6C). Furthermore, when we subsequently prevented both His and Phe at P2 *and* P3 in a new library to suppress potential register shifting, we did not isolate any Ob.1A12- or Ob.2F3-binding peptides (data not shown). These results show that the ‘HF’ motif is required for TCR recognition and its enrichment is a function of TCR preference, not any inherent biases caused by the library or MHC anchor positions of the peptide.

Clustering analysis of the selected peptides for both Ob.1A12 and Ob.2F3 showed distinct clusters consisting of peptides no more than 4 amino acids different from each other (Figure 6D). When the stringency of clustering is increased to allow no more than 3 amino acid differences, matching the analysis done for I-E^k, there were several more sparse clusters (Figure S6E). Since Ob1.A12 and Ob.2F3 are so focused on the HF motif, there are fewer total hotspot residues distributed on the peptide compared to the MCC-reactive TCRs we studied.

High-confidence prediction of naturally occurring TCR-reactive peptides

The surprisingly limited degeneracy of TCR recognition suggests it may be feasible to identify naturally occurring TCR ligands with a random peptide library. However, library selections and deep sequencing alone are not sufficient to identify naturally occurring ligands for two reasons. First, the size of yeast libraries relative to all possible MHC-displayed peptides makes it unlikely that any given peptide sequence exists in the library. Second, the amino acid substitutions that are permitted at each position along the peptide represent a complex, and as our covariation analysis indicated, cooperative interplay between the peptide, MHC, and TCR that may not be well described by common substitution matrices such as BLOSUM (Henikoff and Henikoff, 1992). For example, even

though manual inspection of Ob.1A12-binding sequences readily shows the WT-like 'HF' motif, the sequences do not find MBP as a match in blastp searches (data not shown).

We therefore set out to develop an algorithm to use the aggregate data from our selection results to inform searches for candidate TCR antigens. First, we created a substitution matrix that used the positional frequency information derived from our Ob.1A12 and Ob.2F3 deep sequencing data (Zhao et al., 2001). Since the limited coverage of our libraries could lead to appearance of residue biases at non-critical (i.e. neutral) peptide positions that do not reflect actual selective pressure, we created a new HLA-DR15-based library where we fixed the dominant Ob.1A12 binding motif (P2 His, P3 Phe, and P5 Lys/Arg) along with the P1 and P4 MHC-binding anchors, while randomizing the remaining residues. When the selected libraries were sequenced, we found that while some proximal positions such as P(-1) and P(-2) still showed distinct residue preferences, other positions such as P7 and P8 showed less convergence relative to the original HLA-DR15 library (Figure S7A). The more granular substitution data for peptide positions distal to the TCR-binding 'hotspot' allowed us to construct a more reliable algorithm.

We compiled two 14×20 substitution matrices consisting of the observed frequencies of the 20 amino acids at each of the 14 positions of the library peptides from the focused DR15 pMHC libraries selected by Ob.1A12 and Ob.2F3 (Figure 7A, S7A, Table S2) (Zhao et al., 2001). Since minimal residue covariation was observed for Ob.1A12 and Ob.2F3 selections, each position was treated independently (Figure S7B). Our peptide database search using the Ob.1A12 based matrix yielded 2330 unique hits, including MBP. For the search based on the Ob.2F3 matrix, we had 4824 unique hits, again including MBP. The peptide hits shared the central P(-1)-P5 motif of MBP but the flanking residues showed very little sequence homology to either MBP or to each other (Figure 7B, Table S3). The predicted peptides are from diverse microbial sources, such as bacteria; environmental sources, such as antigens expressed by plants; and from proteins in the human proteome.

To test our computationally predicted ligands for Ob.1A12 and Ob.2F3, we synthesized a diverse set comprising 26 of the potential environmental antigens as well as 7 novel human peptides predicted to cross-react with Ob.1A12 and Ob.2F3. When we tested the 33 putative ligands for activity, 25/26 of the environmental antigens and 6/7 of the human peptides induced proliferation for Ob.1A12 and/or Ob.2F3, a success rate of 94% (Figure 7B).

Discussion

The concept of TCR cross-reactivity is important because key aspects of T cell biology, including thymic development, pathogen surveillance, autoimmunity, and transplant rejection, seemingly require recognition of diverse ligands. In this study, we aimed to define the mechanisms underlying TCR specificity and cross-reactivity using a combinatorial, biochemical approach that yielded massive datasets based on direct selection. This has given us insight into the structural basis of TCR cross-reactivity and also provides a robust way to discover peptide ligands for a TCR of interest.

Our results clarify previous controversies on whether TCRs are highly cross-reactive or highly specific by leveraging large amounts of experimental data found via direct binding of pMHC to TCR. We find that structural principles allow for the TCR to engage large numbers of unique pMHC without requiring degeneracy in pMHC recognition. If the criterion of cross-reactivity is simply the number of unique peptide sequences that can be recognized by any given TCR, then TCRs do exhibit a high degree of cross-reactivity. Given that the libraries greatly undersample all possible sequence combinations it is likely that our hundreds of discovered peptides are indicative of thousands of different peptides can be recognized by the studied TCRs. However, when cross-reactive peptides are examined en masse, we find central conserved TCR-binding (i.e. 'up-facing') motifs. TCR cross-reactivity is not achieved by each receptor recognizing a large number of unrelated peptide epitopes, but rather through greater tolerance for substitutions to peptide residues outside of the TCR interface, differences in residues that contact the MHC, and relatively conservative changes to the residues that contact the TCR CDR loops. The segregation of TCR recognition and MHC binding allow for TCRs to simultaneously accommodate needs for specificity and cross-reactivity, ensuring no 'holes' in the TCR repertoire without requiring degenerate recognition of antigen.

While we believe this mechanism will be general for αβ TCRs, recognition of nonhomologous antigens certainly occurs to varying degrees in the TCR repertoire, although molecularly defined examples are surprisingly rare. The ability for one TCR to bind to multiple MHCs (e.g. alloreactivity), for one TCR to bind in multiple orientations on one MHC, for a peptide to non-canonically bind MHC (e.g. partially-filled MHC grooves, register-shifted peptides), or for a TCR to have TCR-peptide contacts as a disproportionately large or small part of the overall interface (e.g. 'super-bulged' peptides) will grant some receptors a greater degree of epitope promiscuity (Adams et al., 2011; Colf et al., 2007; Maynard et al., 2005; Morris and Allen, 2012; Morris et al., 2011; Tynan et al., 2005). It is also possible that class I versus class II MHC specific TCRs could exhibit different degrees of cross-reactivity as a consequence of the 'low lying' peptides in the class II groove, versus the elevated or 'higher profile' peptides presented by class I. Indeed, in a prior study, multiple peptides reactive with a class I specific (H-2L^{d}) murine TCR were identified through manual curation and the structures indicated a diverse recognition chemistry by the TCR CDR3 loops (Adams et al., 2011). In retrospect, a close inspection reveals striking commonalities in the peptide binding chemistry by the TCR, in particular a requirement for a hydrophobic contact at the apex of the P7 'bulge' that forms the principal site of contact with the TCR CDR3β. In contrast, a second class I TCR, 2C, was not found to exhibit peptide degeneracy, instead exhibiting specificity for its endogenous antigen, QL9, in a manner similar to the class II specific TCRs studied here (unpublished results).

An important implication of our findings, which is consistent with previous studies (Macdonald et al., 2009), is that identification of endogenous antigens of TCRs is feasible using peptide-MHC libraries. In our previous view of cross-reactivity, we assumed that a given TCR would cross-react with so many peptides in a library that elucidation of 'natural' leads from a background of degenerately binding sequences would be extremely difficult. Additionally, the sparse coverage of possible sequences renders it unlikely that any given sequence of interest will be represented with 100% identity in our library. However, limited

TCR epitope cross-reactivity allows us to use selection results to constrain computational searches of protein databases, which proves to be a highly successful strategy to find naturally occurring TCR ligands.. Thus, this approach now opens up the possibility of peptide ligand discovery for 'orphan' TCRs such as those from regulatory T cells and tumor infiltrating lymphocytes (TILs).

While the naturally occurring peptides in this study were found as a proof of principle for our methodology, they further support the hypothesis that autoimmune T cells have the ability to be activated by immunogens encountered in the environment, which may serve as the triggers for the initiation of autoimmunity (De la Herran-Arita et al., 2013; Harkiolaki et al., 2009; Hausmann et al., 1999; Wucherpfennig and Strominger, 1995). Additionally, the potential for other human peptides to cross-react with autoimmune TCRs with previously 'known' antigens presents the intriguing possibility that individual TCRs can recognize multiple self-peptides, potentially contributing to T cell pathologies in autoimmune disease. This notion is supported by the finding that a murine TCR specific for myelin-oligodendrocyte glycoprotein cross-reacts with a second CNS antigen, neurofilament M. Due to this unexpected cross-reactivity, these T cells remained pathogenic even in MOG-deficient mice (Krishnamoorthy et al., 2009). Our approach for systematic discovery of peptides recognized by human TCRs thus has the potential to advance our understanding of complex pathogenesis of immune-mediated diseases.

Methods

Creation and selection of pMHC libraries

Peptide libraries were created through use of mutagenic primers allowing all 20 amino acids via NNK codons. The libraries allowed limited diversity at the known MHC anchor residues to maximize the number of correctly folded and displayed pMHC clones in the library. Yeast libraries were created by electroporation of competent EBY-100 cells via homologous recombination of linearized pYAL vector and mutagenized pMHC construct essentially as described previously (Adams et al., 2011; Chao et al., 2006). Final libraries contained approximately 2×10^8 yeast transformants.

Yeast libraries were selected for binding to TCR of interest coupled to streptavidin coated magnetic beads (Miltenyi) through magnetic activated cell sorting. After libraries enriched above the baseline of streptavidin beads alone (typically after 3 rounds of selection), a final round of selection was conducted with fluorescently labeled streptavidin tetramers.

Deep sequencing of pMHC libraries

Pooled plasmids from 5×10^7 yeast from each round of selection were isolated via yeast miniprep (Zymoprep II kit, Zymo Research) and used as PCR template to prepare sequencing samples. The adapter and barcode sequences were appended via nested 25-round cycles of PCR of the purified plasmids using Phusion polymerase (NEB). Deep sequencing was conducted on an Illumina MiSeq sequencer at the Stanford Stem Cell Institute Genome Center.

Profile-based searches for naturally occurring peptide ligands

The positional frequencies from round 3 of the fixed HF library were used to generate a 14×20 substitution matrix. Each protein in the NR (NCBI) or human protein (Uniprot) databases was scanned using a 14 position sliding window and scored as a product of the positional substitution matrix (Cockcroft and Osguthorpe, 1991). In this way, a candidate peptide containing even a single disallowed substitution would be excluded as a possible hit.

Structural determination of pMHC-TCR complexes

All crystallographic data was collected at the Stanford Synchroton Radiation Lightsource (Stanford, CA) beamlines 11-1 and 12-2. Data were indexed, integrated, and scaled using either the XDS or the HKL-2000 program suites (Kabsch, 2010; Otwinowski et al., 1997). All structures were solved via molecular replacement using the program Phaser (McCoy, 2007) and refined with Phenix (Adams et al., 2010).

Extended experimental procedures

Further details for the design, selection, and sequencing of yeast display libraries; methods for production, characterization, and crystallization of proteins; and computational discovery and functional validation of peptide hits can be found in the extended experimental procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Suzanne Fischer, Gary Mantalas, and Nelida Prado for technical assistance, and Jarrett Adams, Lauren Ely, and Evan Newell for helpful discussions. M.E.B was supported by a Regina Casper Stanford Graduate Fellowship, a Gerald J. Lieberman Fellowship, and a National Science Foundation Graduate Fellowship. J.L.M was supported by NCIK01CA175127 and a Helen Hay Whitney Foundation postdoctoral fellowship. This work was supported by the NIH (PO1 AI045757 to K.W.W., and R01 AI03867 and U19 4100041120 to K.C.G.), and HHMI (to M.M.D and K.C.G.).

References

Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, Chen W, Levin AM, Connolly JM, Zhu C, et al. T Cell Receptor Signaling Is Limited by Docking Geometry to Peptide-Major Histocompatibility Complex. Immunity. 2011; 35:681–693. [PubMed: 22101157]

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

Basu D, Horvath S, Matsumoto I, Fremont DH, Allen PM. Molecular basis for recognition of an arthritic peptide and a foreign epitope on distinct MHC molecules by a single TCR. J Immunol. 2000; 164:5788–5796. [PubMed: 10820257]

Benoist C, Mathis D. Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? Nat Immunol. 2001; 2:797–801. [PubMed: 11526389]

Birnbaum ME, Dong S, Garcia KC. Diversity-oriented approaches for interrogating T-cell receptor repertoire, ligand recognition, and function. Immunol Rev. 2012; 250:82–101. [PubMed: 23046124]

Boder ET, Bill JR, Nields AW, Marrack PC, Kappler JW. Yeast surface display of a noncovalent MHC class II heterodimer complexed with antigenic peptide. Biotechnol Bioeng. 2005; 92:485–491. [PubMed: 16155952]

Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc. 2006; 1:755–768. [PubMed: 17406305]

Cockcroft VB, Osguthorpe DJ. Relative-residue surface-accessibility patterns reveal myoglobin and catalase similarity. FEBS letters. 1991; 293:149–152. [PubMed: 1959649]

Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, Kranz DM, Garcia KC. How a single T cell receptor recognizes both self and foreign MHC. Cell. 2007; 129:135–146. [PubMed: 17418792]

Crawford F, Huseby E, White J, Marrack P, Kappler JW. Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. PLoS Biol. 2004; 2:E90. [PubMed: 15094798]

Crawford F, Jordan KR, Stadinski B, Wang Y, Huseby E, Marrack P, Slansky JE, Kappler JW. Use of baculovirus MHC/peptide display libraries to characterize T-cell receptor ligands. Immunol Rev. 2006; 210:156–170. [PubMed: 16623770]

Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. Nature. 1988; 334:395–402. [PubMed: 3043226]

De la Herran-Arita AK, Kornum BR, Mahlios J, Jiang W, Lin L, Hou T, Macaubas C, Einen M, Plazzi G, Crowe C, et al. CD4+ T Cell Autoimmunity to Hypocretin/Orexin and Cross-Reactivity to a 2009 H1N1 Influenza A Epitope in Narcolepsy. Science translational medicine. 2013; 5:216ra176.

Ebert PJ, Jiang S, Xie J, Li QJ, Davis MM. An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nat Immunol. 2009; 10:1162–1169. [PubMed: 19801983]

Ehrich EW, Devaux B, Rock EP, Jorgensen JL, Davis MM, Chien YH. T cell receptor interaction with peptide/major histocompatibility complex (MHC) and superantigen/MHC ligands is dominated by antigen. J Exp Med. 1993; 178:713–722. [PubMed: 8393480]

Esteban O, Zhao H. Directed evolution of soluble single-chain human class II MHC molecules. J Mol Biol. 2004; 340:81–95. [PubMed: 15184024]

Evavold BD, Sloan-Lancaster J, Wilson KJ, Rothbard JB, Allen PM. Specific T cell recognition of minimally homologous peptides: evidence for multiple endogenous ligands. Immunity. 1995; 2:655–663. [PubMed: 7540944]

Garcia KC, Adams EJ. How the T cell receptor sees antigen--a structural view. Cell. 2005; 122:333–336. [PubMed: 16096054]

Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. 2007; 357:851–862. [PubMed: 17660530]

Hahn M, Nicholson MJ, Pyrdol J, Wucherpfennig KW. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. Nat Immunol. 2005; 6:490–496. [PubMed: 15821740]

Harkiolaki M, Holmes SL, Svendsen P, Gregersen JW, Jensen LT, McMahon R, Friese MA, van Boxel G, Etzensperger R, Tzartos JS, et al. T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. Immunity. 2009; 30:348–357. [PubMed: 19303388]

Hausmann S, Martin M, Gauthier L, Wucherpfennig KW. Structural features of autoreactive TCR that determine the degree of degeneracy in peptide recognition. J Immunol. 1999; 162:338–344. [PubMed: 9886404]

Hemmer B, Fleckenstein BT, Vergelli M, Jung G, McFarland H, Martin R, Wiesmuller KH. Identification of high potency microbial and self ligands for a human autoreactive class II-restricted T cell clone [see comments]. J Exp Med. 1997; 185:1651–1659. [PubMed: 9151902]

Hemmer B, Vergelli M, Gran B, Ling N, Conlon P, Pinilla C, Houghten R, McFarland HF, Martin R. Predictable TCR antigen recognition based on peptide scans leads to the identification of agonist ligands with no sequence homology. J Immunol. 1998a; 160:3631–3636. [PubMed: 9558061]

Hemmer B, Vergelli M, Pinilla C, Houghten R, Martin R. Probing degeneracy in T-cell recognition using peptide combinatorial libraries. Immunol Today. 1998b; 19:163–168. [PubMed: 9577092]

Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992; 89:10915–10919. [PubMed: 1438297]

Jiang W, Boder ET. High-throughput engineering and analysis of peptide binding to class II MHC. Proc Natl Acad Sci U S A. 2010; 107:13258–13263. [PubMed: 20622157]

Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]

Kersh GJ, Allen PM. Essential flexibility in the T-cell recognition of antigen. Nature. 1996; 380:495–498. [PubMed: 8606766]

Krishnamoorthy G, Saxena A, Mars LT, Domingues HS, Mentele R, Ben-Nun A, Lassmann H, Dornmair K, Kurschus FC, Liblau RS, et al. Myelin-specific T cells also recognize neuronal autoantigen in a transgenic mouse model of multiple sclerosis. Nature medicine. 2009; 15:626–632.

Krogsgaard M, Prado N, Adams EJ, He XL, Chow DC, Wilson DB, Garcia KC, Davis MM. Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation. Mol Cell. 2003; 12:1367–1378. [PubMed: 14690592]

Lo WL, Felix NJ, Walters JJ, Rohrs H, Gross ML, Allen PM. An endogenous peptide positively selects and augments the activation and survival of peripheral CD4+ T cells. Nat Immunol. 2009; 10:1155–1161. [PubMed: 19801984]

Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, Clements CS, Bharadwaj M, Kjer-Nielsen L, Saunders PM, Wilce MC, et al. T cell allorecognition via molecular mimicry. Immunity. 2009; 31:897–908. [PubMed: 20064448]

Madsen LS, Andersson EC, Jansson L, krogsgaard M, Andersen CB, Engberg J, Strominger JL, Svejgaard A, Hjorth JP, Holmdahl R, et al. A humanized model for multiple sclerosis using HLA-DR2 and a human T-cell receptor. Nature genetics. 1999; 23:343–347. [PubMed: 10610182]

Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunol Today. 1998; 19:395–404. [PubMed: 9745202]

Matsui K, Boniface JJ, Reay PA, Schild H, Fazekas de St Groth B, Davis MM. Low affinity interaction of peptide-MHC complexes with T cell receptors. Science. 1991; 254:1788–1791. [PubMed: 1763329]

Maynard J, Petersson K, Wilson DH, Adams EJ, Blondelle SE, Boulanger MJ, Wilson DB, Garcia KC. Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. Immunity. 2005; 22:81–92. [PubMed: 15664161]

Mazza C, Auphan-Anezin N, Gregoire C, Guimezanes A, Kellenberger C, Roussel A, Kearney A, van der Merwe PA, Schmitt-Verhulst AM, Malissen B. How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? EMBO Journal. 2007; 26:1972–1983. [PubMed: 17363906]

McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr. 2007; 63:32–41. [PubMed: 17164524]

Morris GP, Allen PM. How the TCR balances sensitivity and specificity for the recognition of self and pathogens. Nat Immunol. 2012; 13:121–128. [PubMed: 22261968]

Morris GP, Ni PP, Allen PM. Alloreactivity is limited by the endogenous peptide repertoire. Proc Natl Acad Sci U S A. 2011; 108:3695–3700. [PubMed: 21321209]

Nanda NK, Arzoo KK, Geysen HM, Sette A, Sercarz EE. Recognition of multiple peptide cores by a single T cell receptor. J Exp Med. 1995; 182:531–539. [PubMed: 7629510]

Newell EW, Ely LK, Kruse AC, Reay PA, Rodriguez SN, Lin AE, Kuhns MS, Garcia KC, Davis MM. Structural Basis of Specificity and Cross-Reactivity in T Cell Receptors Specific for Cytochrome c-I-E. Journal of immunology. 2011; 186:5823–5832.

Otwinowski, Z.; Minor, W.; Charles, W. Carter, Jr Methods in Enzymology. Academic Press; 1997. [20] Processing of X-ray diffraction data collected in oscillation mode; p. 307-326.

Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, Raj T, Imsgc, Anzgene, Gourraud PA, et al. Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. PLoS genetics. 2013; 9:e1003926. [PubMed: 24278027]

Reay PA, Kantor RM, Davis MM. Use of global amino acid replacements to define the requirements for MHC binding and T cell recognition of moth cytochrome c (93-103). J Immunol. 1994; 152:3946–3957. [PubMed: 7511662]

Reiser JB, Darnault C, Gregoire C, Mosser T, Mazza G, Kearney A, van der Merwe PA, Fontecilla-Camps JC, Housset D, Malissen B. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. Nature Immunology. 2003; 4:241–247. [PubMed: 12563259]

Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annual Review of Immunology. 2006; 24:419–466.

Shann F, Nohynek H, Scott JA, Hesseling A, Flanagan KL. Randomized trials to study the nonspecific effects of vaccines in children in low-income countries. Pediatr Infect Dis J. 2010; 29:457–461. [PubMed: 20431383]

Shih FF, Allen PM. T cells are not as degenerate as you think, once you get to know them. Mol Immunol. 2004; 40:1041–1046. [PubMed: 15036908]

Starwalt SE, Masteller EL, Bluestone JA, Kranz DM. Directed evolution of a single-chain class II MHC product by yeast display. Protein engineering. 2003; 16:147–156. [PubMed: 12676983]

Tynan FE, Burrows SR, Buckle AM, Clements CS, Borg NA, Miles JJ, Beddoe T, Whisstock JC, Wilce MC, Silins SL, et al. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. Nature Immunology. 2005; 6:1114–1122. [PubMed: 16186824]

Wang Y, Rubtsov A, Heiser R, White J, Crawford F, Marrack P, Kappler JW. Using a baculovirus display library to identify MHC class I mimotopes. Proc Natl Acad Sci U S A. 2005; 102:2476–2481. [PubMed: 15699351]

Welsh RM, Che JW, Brehm MA, Selin LK. Heterologous immunity between viruses. Immunol Rev. 2010; 235:244–266. [PubMed: 20536568]

Wen F, Esteban O, Zhao H. Rapid identification of CD4+ T-cell epitopes using yeast displaying pathogen-derived peptide library. J Immunol Methods. 2008; 336:37–44. [PubMed: 18448115]

Wen F, Sethi DK, Wucherpfennig KW, Zhao H. Cell surface display of functional human MHC class II proteins: yeast display versus insect cell display. Protein Eng Des Sel. 2011; 24:701–709. [PubMed: 21752831]

Wilson DB, Pinilla C, Wilson DH, Schroder K, Boggiano C, Judkowski V, Kaye J, Hemmer B, Martin R, Houghten RA. Immunogenicity. I. Use of peptide libraries to identify epitopes that activate clonotypic CD4+ T cells and induce T cell responses to native peptide ligands. J Immunol. 1999; 163:6424–6434. [PubMed: 10586032]

Wilson DB, Wilson DH, Schroder K, Pinilla C, Blondelle S, Houghten RA, Garcia KC. Specificity and degeneracy of T cells. Mol Immunol. 2004; 40:1047–1055. [PubMed: 15036909]

Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, Dolton G, Clement M, Llewellyn-Lacey S, Price DA, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. J Biol Chem. 2012; 287:1168–1177. [PubMed: 22102287]

Wu LC, Tuot DS, Lyons DS, Garcia KC, Davis MM. Two-step binding mechanism for T-cell receptor recognition of peptide MHC. Nature. 2002; 418:552–556. [PubMed: 12152083]

Wucherpfennig KW, Allen PM, Celada F, Cohen IR, De Boer R, Garcia KC, Goldstein B, Greenspan R, Hafler D, Hodgkin P, et al. Polyspecificity of T cell and B cell receptor recognition. Semin Immunol. 2007; 19:216–224. [PubMed: 17398114]

Wucherpfennig KW, Sette A, Southwood S, Oseroff C, Matsui M, Strominger JL, Hafler DA. Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. J Exp Med. 1994a; 179:279–290. [PubMed: 7505801]

Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. Cell. 1995; 80:695–705. [PubMed: 7534214]

Wucherpfennig KW, Zhang J, Witek C, Matsui M, Modabber Y, Ota K, Hafler DA. Clonal expansion and persistence of human T cells specific for an immunodominant myelin basic protein peptide. J Immunol. 1994b; 152:5581–5592. [PubMed: 7514641]

Yin Y, Mariuzza RA. The multiple mechanisms of T cell receptor cross-reactivity. Immunity. 2009; 31:849–851. [PubMed: 20064442]

Zhao R, Loftus DJ, Appella E, Collins EJ. Structural evidence of T cell xeno-reactivity in the absence of molecular mimicry. J Exp Med. 1999; 189:359–370. [PubMed: 9892618]

Zhao Y, Gran B, Pinilla C, Markovic-Plese S, Hemmer B, Tzou A, Whitney LW, Biddison WE, Martin R, Simon R. Combinatorial peptide libraries and biometric score matrices permit the quantitative analysis of specific and degenerate interactions between clonotypic TCR and MHC peptide ligands. J Immunol. 2001; 167:2130–2141. [PubMed: 11489997]

Highlights

- Deep sequencing peptide-MHC libraries finds hundreds of TCR-reactive peptides
- TCRs exhibit limited cross-reactivity for contact residues in peptide antigens
- Structures show linkages between distantly related peptide sequences
- Novel strategy for identification of naturally occurring TCR ligands

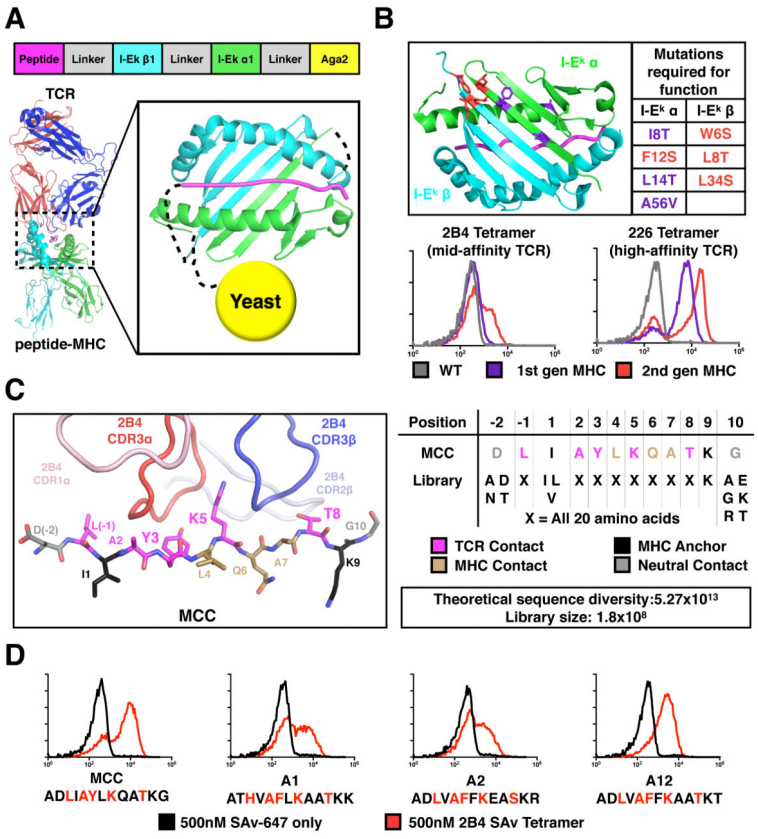

Figure 1. Library design and selection of I-E^k, a murine class II MHC molecule

(A) Schematic of the murine class II MHC I-E^k displayed on yeast, as β1α1 'mini' MHC with peptide covalently linked to MHC N-terminus. (B) Mutations required for correct folding of the β1α1 'mini' I-E^k (top). Mutations were derived from error prone mutagenesis and selection (purple) and rational design (red). Staining with 2B4 and 226 tetramers demonstrate function of error prone-only construct (1st gen MHC) as well as error prone + designed mutant construct (2nd gen MHC) (bottom). (C) Design of the peptide library displayed by I-E^k. Design is based upon the structure of 2B4 bound to MCC-I-E^k (left). Residues from P(-2) to P10 are randomized, with limited diversity at P(-2), P10, and the P1/P9 anchors (right). (D) 500 nM TCR tetramer staining of three clones selected for binding to 2B4 TCR compared to MCC (wild-type). TCR contact residues are colored red. See also Figure S1.

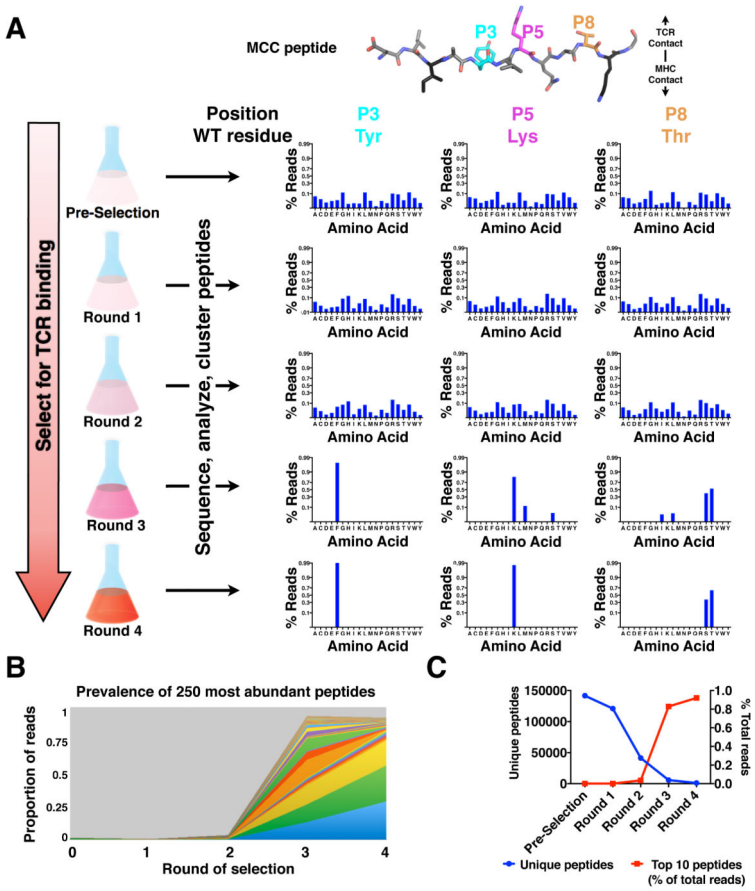

Figure 2. Deep sequencing of peptide selections on I-E^{k} converges on one dominant epitope for 2B4 TCR recognition

(A) Plots for amino acid prevalence at the three primary TCR contact positions (P3 (cyan), P5 (magenta), and P8 (orange)) show the peptide library enriches from even representation of all amino acids in the pre-selection library to a WT-like motif at each position. A secondary preference can be seen at P5 and P8 in round 3 but is outcompeted by round 4. (B) Sequence enrichment of 250 most abundant peptides show a convergence from a broad array of sequences to a few related clones. Area in grey represents all clones other than the most prevalent 250. (C) Comparison of total number of peptides and prevalence of 10 most abundant peptides for each round of selection. See also Figure S2.

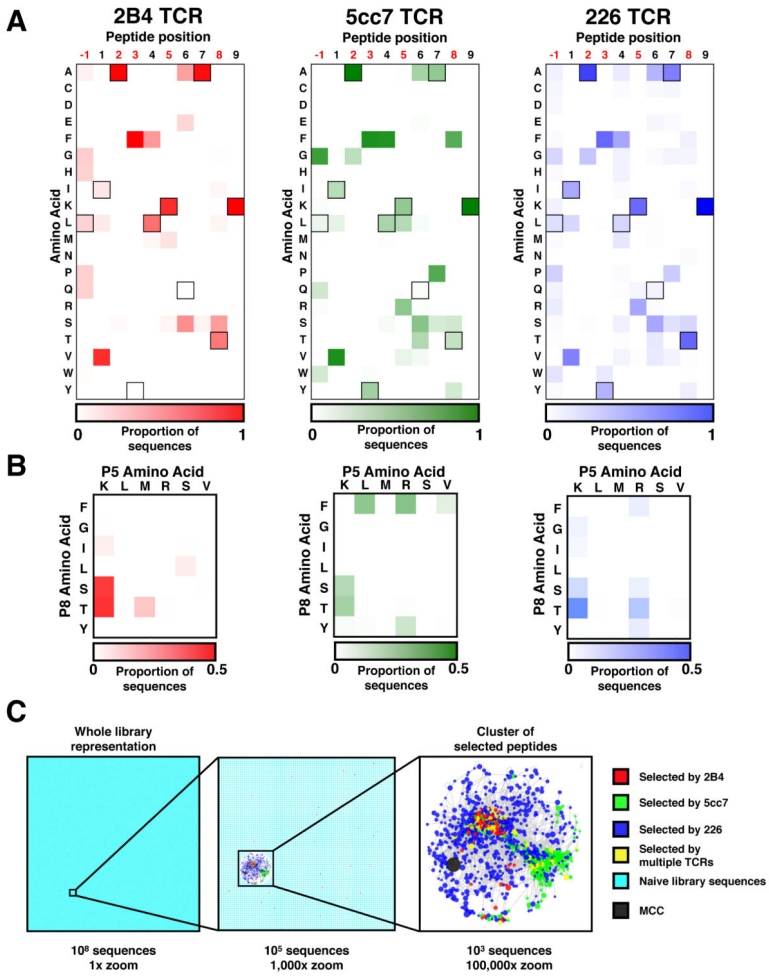

Figure 3. Three different MCC/I-E^{k} reactive TCRs require a WT-like recognition motif in the peptide antigens

(A) Heatmaps of amino acid preference by position for 2B4 (left, red) 5cc7 (center, green) and 226 (right, blue) TCRs after three rounds of selection. The sequence for MCC is represented via outlined boxes. TCR contact residues are labeled red on x axis. (B) Covariation analysis of TCR contact positions P5 (x axis) and P8 (y axis) show distinct coupling of amino acid preferences. (C) Minimum distance clustering of all TCR sequences selected above background show sequences for all TCRs form one large cluster with MCC (black circle, not represented in library but added for reference). Sequence cluster placed in a representation of whole-library sequence space (left: 1x magnification, center: 1000x magnification) for reference. See also Figure S3.

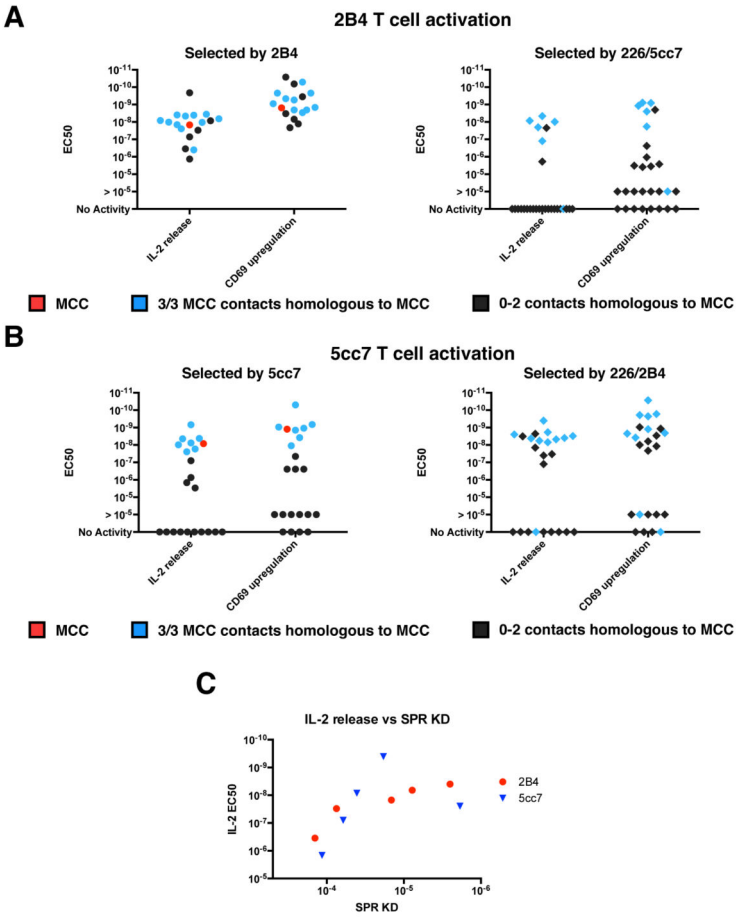

Figure 4. Relationships between affinity and activity of peptides selected for binding to IE^{k}-reactive TCRs

(A) EC50s of IL-2 release and CD69 upregulation for 2B4 T cells stimulated with peptides selected with 2B4 TCR, plus MCC (red) (left), or peptides selected with 226 or 5cc7 TCRs (right). Sequences with close homology to MCC at P3, P5, and P8 are represented in blue. Sequences that do not share 3/3 TCR contacts with MCC are in black. (B) EC50s as in A for 5cc7 T cells with peptides selected with 5cc7 (left) or 226/2B4 (right) TCRs. (C) Correlation between TCR-pMHC affinity and peptide signaling potency. Each data point represents one peptide. See also Figure S4.

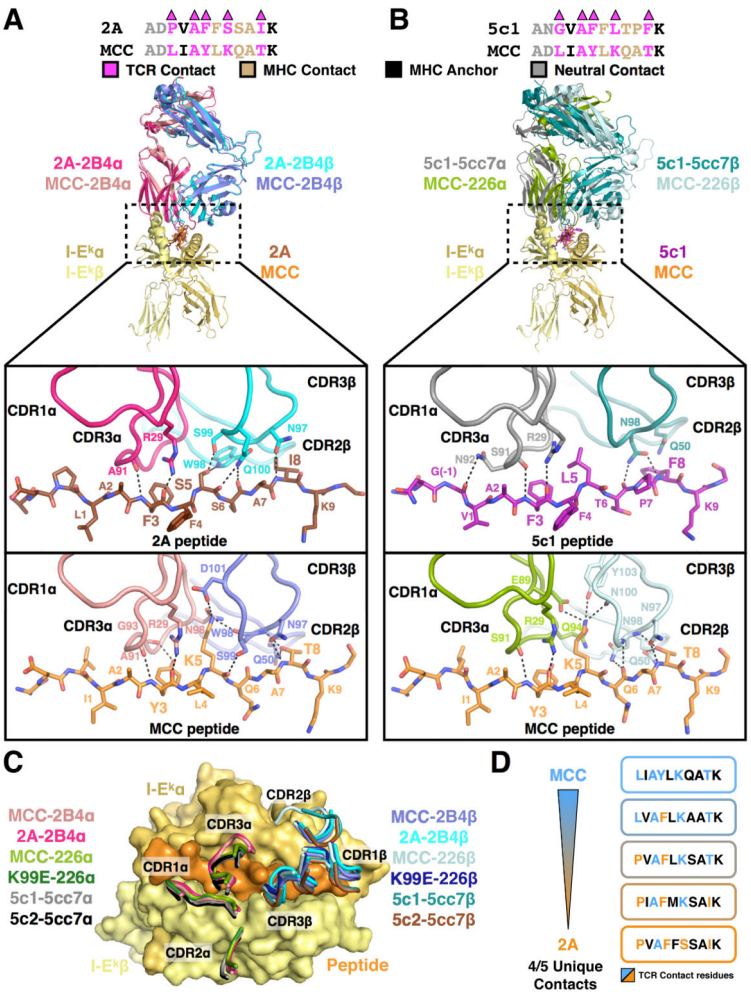

Figure 5. Peptides distantly related to MCC show highly similar mechanism of recognition and linkages to the cognate antigen

Crystal structures of TCR-pMHC complexes for 2B4-2A-I-E^k and 2B4-MCC-I-E^k (PDB ID: 3QIB) (A) and 5cc7-5c1-I-E^k and 226-MCC-I-E^k (PDB ID: 3QIU) (B) compared. TCR contacts are shown in magenta (top, noted with triangles). There is very little change in overall binding geometry despite significant variation of peptide sequence. The TCRs accommodate differences in peptide sequence primarily through differences in CDR3β (bottom). (C) TCR CDR loop footprints for 2B4 recognizing MCC and 2A peptides, 226 recognizing MCC and MCC K99E peptides, and 5cc7 recognizing 5c1 and 5c2 peptides show very little deviation. (D) Relationship between MCC and 2A peptides revealed through intermediate selected peptide sequences. See also Table S1 and Figure S5.

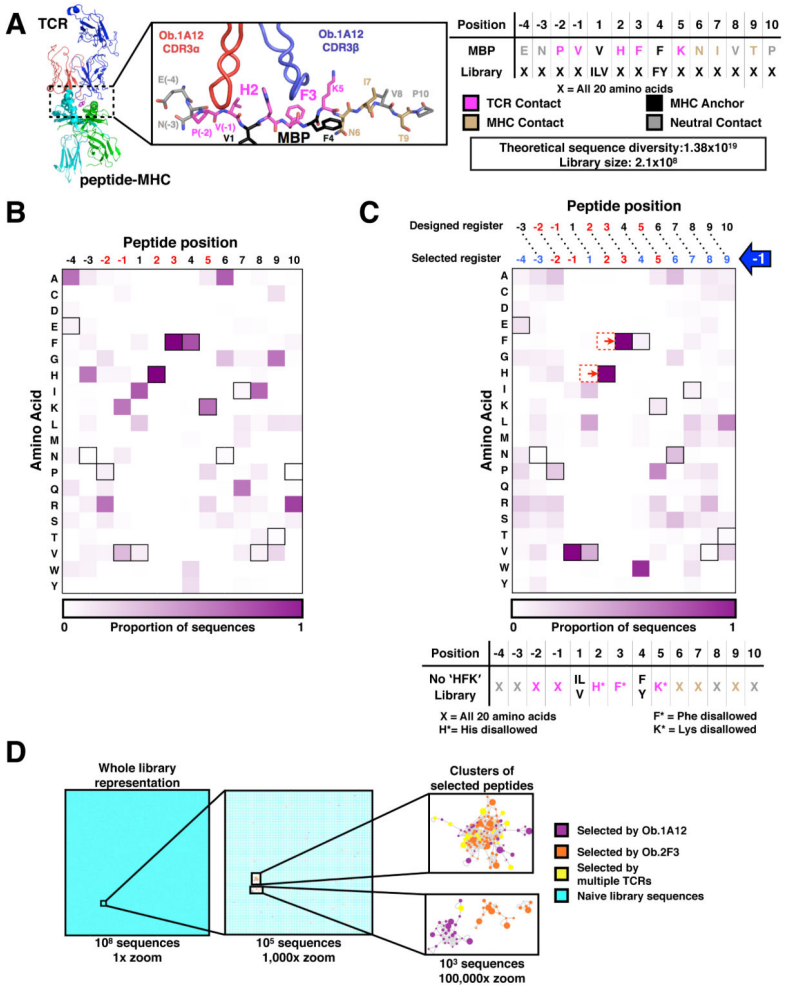

Figure 6. Design and selection of HLA-DR15 based libraries for myelin basic protein (MBP)-reactive human TCRs

(A) HLA-DR15 library design based upon structure of Ob.1A12-MBP-HLA-DR15 complex. Residues P(-4)-P10 are fully randomized, except for the P1 and P4 anchors (in black). TCR contacts are colored magenta. (B) Heatmap of amino acid preference by position for Ob.1A12 TCR. The sequence for MBP is represented via outlined boxes. TCR contacts are labeled red on the x axis. (C) Design and selection results of library that suppresses central 'HF' TCR recognition motif at P2-P3 of peptide. Resulting register shift is shown in blue on x axis. (D) Sequence clustering shows distinct, related clusters of selected peptides. Sequence cluster placed in a representation of whole-library sequence space (left: 1x magnification, center: 1000x magnification) for reference. See also Figure S6.

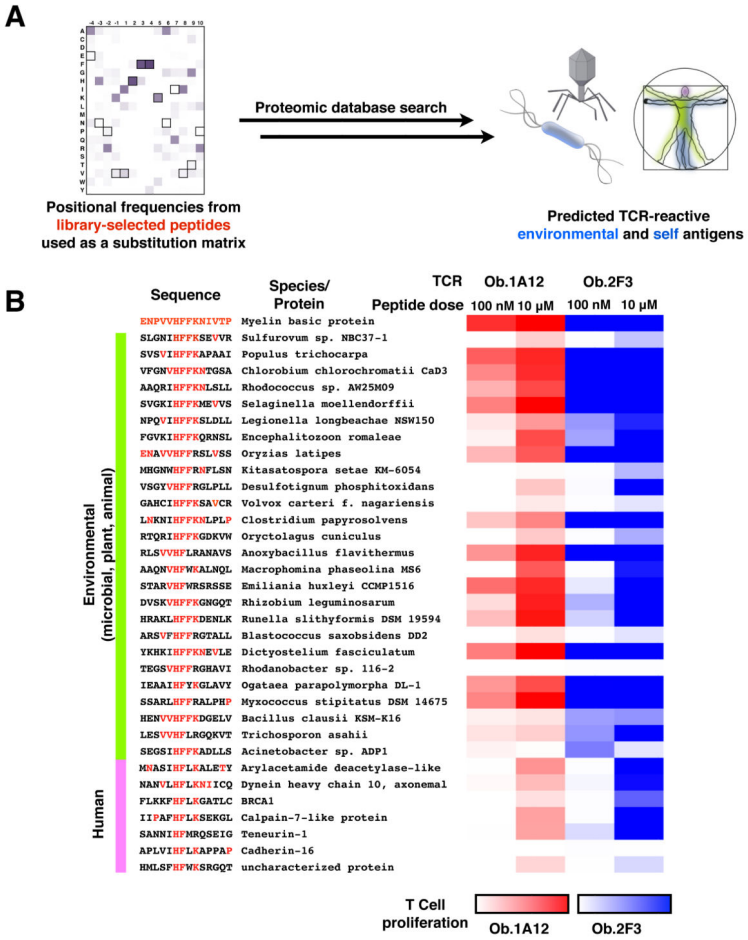

Figure 7. Discovery of naturally occurring TCR ligands through deep sequencing and substitution matrix-based homology search

(A) Schematic for ligand search strategy, in which a positional substitution matrix is generated from deep sequencing data and then used to find naturally occurring peptides that are represented within the matrix. (B) Functional characterization of a selection of naturally occurring peptides with predicted activity. Activity is tested via proliferation of T cells when exposed to peptide. Heatmaps are normalized to 10μM dose of MBP peptide for each T cell clone. See also Tables S2 and S3 and Figure S7.