

Understanding the drivers of MHC restriction of T cell receptors

Nicole L. La Gruta^{1,2*}, Stephanie Gras^{1,2}, Stephen R. Daley¹, Paul G. Thomas³ and Jamie Rosjohn^{1,2,4*}

Abstract | T cell discrimination of self and non-self is predicated on $\alpha\beta$ T cell receptor (TCR) co-recognition of peptides presented by MHC molecules. Over the past 20 years, structurally focused investigations into this MHC-restricted response have provided profound insights into T cell function. Simultaneously, two models of TCR recognition have emerged, centred on whether the TCR has, through evolution, acquired an intrinsic germline-encoded capacity for MHC recognition or whether MHC reactivity is conferred by developmental selection of TCRs. Here, we review the structural and functional data that pertain to these theories of TCR recognition, which indicate that it will be necessary to assimilate features of both models to fully account for the molecular drivers of this evolutionarily ancient interaction between the TCR and MHC molecules.

In 1974, Peter Doherty and Rolf Zinkernagel discovered that T cell activation requires simultaneous co-recognition of fragments of foreign peptide antigen and self MHC molecules¹. This seminal finding, and the subsequent discovery of the $\alpha\beta$ T cell receptor (TCR) that is responsible for mediating this recognition^{2,3}, revealed a receptor–ligand interaction system that is essentially unparalleled in biology — namely, the combined need for a given TCR to recognize both a self MHC molecule and a diverse array of self-derived and pathogen-derived peptides. The delicate balance of this unlikely equilibrium is the basis of T cell-mediated immunity, which provides effective protection from infection while preventing T cell-mediated autoimmunity.

There has been continuous progress in our understanding of TCR recognition and signalling events over the past two decades, which has been advanced by many technological developments^{4–6} (FIG. 1). Nevertheless, there remains a central controversy regarding the molecular drivers of the interaction between the TCR and peptide–MHC (pMHC). Two theories have been proposed to explain how TCR recognition of pMHC is specified: the germline-encoded theory^{7–11}, which is based on Niels Jerne's theory of an evolutionary 'hardwiring' of the TCR for recognition of MHC molecules through germline-encoded motifs¹², and the selection theory of TCR recognition^{13–17}, which suggests that extreme randomness of TCR diversity has been maintained during evolution and that TCR editing during development imposes the constraint of MHC recognition.

Following the twentieth anniversary of the Nobel Prize in Physiology or Medicine being awarded to Zinkernagel and Doherty in 1996 for the landmark

discovery of MHC restriction of TCRs, it is pertinent to revisit our understanding of how the adaptive immune system has solved the complex biological problem of simultaneous self and non-self recognition. In doing so, we advance our knowledge of the fundamentals of adaptive immunity and its evolution. A compounding factor is that T cells are not activated solely by TCR–pMHC recognition. Rather, a large number of co-receptors and accessory co-stimulatory molecules, as well as the CD3 signalling machinery, collectively determine whether a T cell activation signal is elicited⁶. Thus, TCR–pMHC recognition leads to T cell activation through a multifactorial process that is complicated by the extreme diversity inherent within this system. Here, we review the two theories that have been proposed to explain TCR recognition of MHC (BOX 1), discuss the implications of each for T cell development and signalling and propose an amalgamation of these models on the basis of the available structural and functional evidence.

MHC molecules and TCRs: a numbers game

An infinite number of peptide-based ligands could potentially arise from the array of proteins that are encoded by a host and its pathogens, which is further increased by various forms of post-translational modification^{18,19}. Moreover, TCRs can also interact with lipid and metabolite-derived antigens when presented by MHC class I-like molecules (reviewed in REFS^{20–22}). To cope with this diversity of potential antigens, the immune system has developed a system for antigen display and recognition based on MHC molecules and TCRs, respectively.

¹Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia.

²ARC Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria, Australia.

³Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN, USA.

⁴Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK.

*e-mail: nicole.la.gruta@monash.edu; jamie.rosjohn@monash.edu

<https://doi.org/10.1038/s41577-018-0007-5>

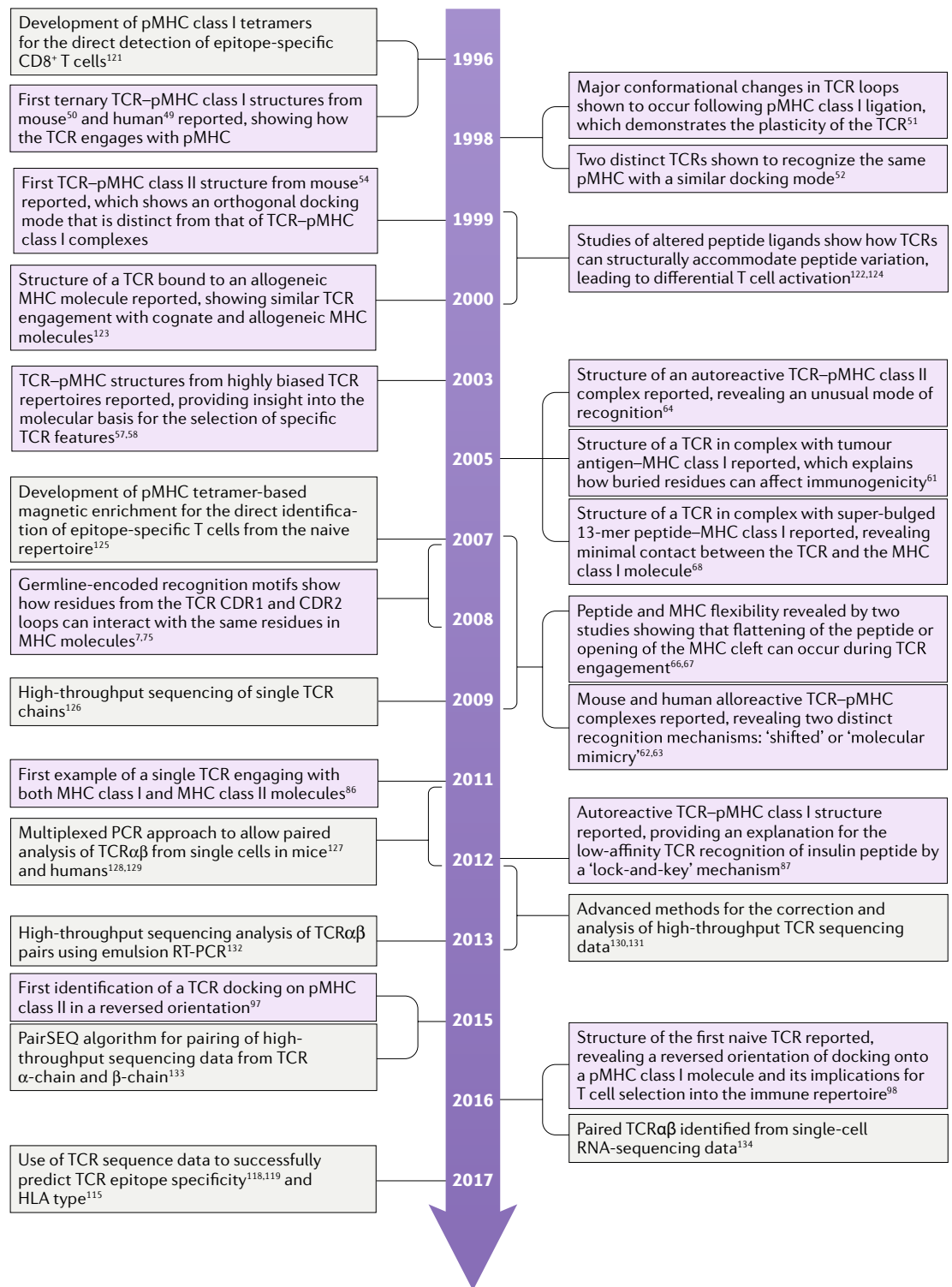


Fig. 1 | **Chronological view of technological developments and conceptual advances that have furthered our understanding of TCR recognition of peptide–MHC.** Purple text boxes indicate conceptual advances; grey text boxes indicate technological advances. CDR, complementarity-determining region; pMHC, peptide–MHC; RT-PCR, reverse transcription PCR; TCR, T cell receptor.

MHC molecules. Structural studies have shown how MHC molecules, which are subdivided into two classes (MHC class I and MHC class II), capture peptides. MHC class I molecules are composed of a heavy chain and light chain (β_2 -microglobulin), with the antigen-binding cleft

within the heavy chain being composed of two α -helical ‘jaws’ and a β -sheet floor²³ (FIG. 2a). The peptide is bound within this antigen-binding cleft, which is pinched-off at the termini and thereby generally favours the binding of peptides of 8–10 amino acids in length. Nevertheless,

Box 1 | Models for MHC restriction of TCR recognition

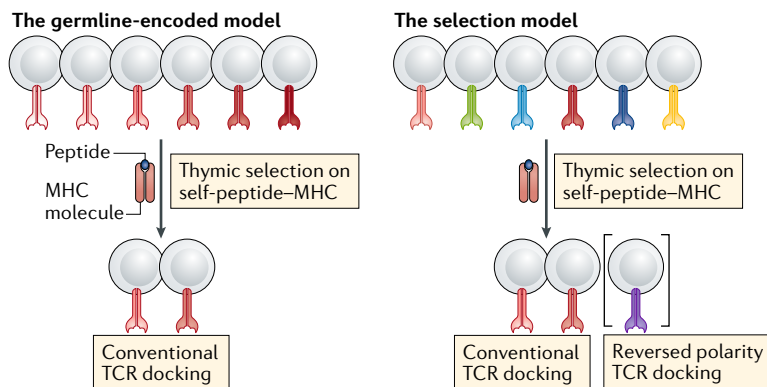
Two models have been proposed to explain the mechanistic basis of T cell receptor (TCR) recognition of MHC molecules (see the figure).

The germline-encoded model

This model, which is based on Niels Jerne's original hypothesis¹², posits that TCR genes have, through millions of years of co-evolution with the MHC, undergone selection for intrinsic recognition of MHC molecules (TCRs with an ability to recognize MHC molecules are depicted in shades of red). Consequently, pairwise interactions between evolutionarily conserved amino acid residues encoded by TCR V gene elements and MHC genes are thought to drive the preferential association between TCRs and MHC molecules^{7–11}. The model proposes that multiple such 'interaction codons' exist that are unique to particular TCR V gene–MHC combinations. Support for this model comes predominantly from studies that have shown evolutionary conservation of TCR residues that commonly contact MHC molecules, most notably Tyr48 in complementarity-determining region 2 of the TCR β -chain (CDR2 β)¹⁰, as well as the prevalence (up to 30%) of intrinsic MHC reactivity in pre-selection TCR repertoires^{33,34,36}. This model suggests that non-MHC restriction of TCRs or reversed polarity MHC recognition by TCRs occurs rarely and inadvertently as a consequence of stochastic cross reactivity.

The selection model

This model proposes that the TCR has no intrinsic reactivity to MHC molecules (as depicted by the TCRs of various colours) but that MHC reactivity (as indicated by the red TCRs) is conferred by signalling constraints imposed during thymic positive selection^{13–17}. Specifically, it proposes that focusing of the TCR repertoire onto the recognition of MHC molecules is driven by the requirement for CD4 or CD8 co-receptor binding to an MHC molecule for efficient localization of the tyrosine-protein kinase LCK to the CD3 signalling complex. Support for this model derives from the large number of non-MHC-restricted TCRs in the pre-selection TCR repertoire^{14,16} as well as, collectively, from several examples of unconventional recognition by TCRs in the naive T cell repertoire, including reversed polarity docking of the TCR onto the peptide–MHC^{97,98} and MHC-independent TCR recognition of antigen^{104–108}. This model supports unorthodox modes of TCR recognition of MHC molecules (as indicated by the purple TCR) so long as signalling capacity is maintained.



approximately 5–10% of bound peptides are longer, typically protruding outside of the MHC class I antigen-binding cleft²⁴. MHC class II molecules are composed of an α -chain and a β -chain that form an antigen-binding cleft analogous to that of MHC class I molecules. However, the open-ended nature of the MHC class II cleft enables peptides of greater length (more than 14 amino acids) to bind and be presented for TCR recognition²⁵ (reviewed in REFS^{4,26}) (FIG. 2b).

Pockets within the antigen-binding cleft of a given MHC molecule determine its peptide-binding preferences and register, which are in turn shaped by MHC polymorphisms²⁶. Indeed, the MHC locus is the most polymorphic region of the human genome, with more than 6,000 MHC molecules having been described so far. Such polymorphism enables MHC molecules to

present a diverse array of peptide antigens, with different MHC allomorphs having distinct peptide-binding preferences that are determined by anchor residues that reside within certain MHC pockets. For example, the P1 and P9 pockets of HLA-DQ8 are ideally suited to accommodate glutamate²⁷, whereas proline and aromatic residues are preferentially bound within the P2 and P Ω pockets, respectively, of HLA-B35 (REFS^{28,29}). Thus, there are a large number of pMHC 'barcodes' that need to be efficiently scanned by T cells.

TCRs. The scanning function of $\alpha\beta$ T cells is accomplished by the TCR, which comprises two chains (α and β), each of which is made up of several gene segments (α -chain: *TRAV* and *TRAJ*; β -chain: *TRBV*, *TRBD* and *TRBJ*) as well as non-templated nucleotide (N) additions and deletions at gene junctional boundaries (FIG. 2c). The recognition site for pMHC is typically formed from the complementarity-determining region (CDR) loops (three from each TCR chain). The CDR1 and CDR2 loops are germline encoded by the *TRAV* or *TRBV* genes, whereas the CDR3 loops are generated from the V-(N)-(D)-(N)-J gene junctions and thereby have greater diversity than the CDR1 and CDR2 loops^{4,30}. In humans, the TCR α -chain locus comprises 47 *TRAV* genes and 61 *TRAJ* genes, and the TCR β -chain locus contains 54 *TRBV*, 2 *TRBD* and 14 *TRBJ* genes³¹. Theoretically, this gives rise to 10^{15} – 10^{21} potential TCRs, which provides the diversity that defines adaptive immunity³². A challenge is to understand the molecular rules that govern the TCR–pMHC interaction against this backdrop of extraordinary diversity.

In the following sections, we discuss the evidence in support of the germline-encoded and selection theories of MHC restriction in the context of studies of the pre-selection TCR repertoire, TCR–pMHC structural studies and the requirements for effective TCR signalling. Finally, we outline how the rapidly evolving field of systems immunology has facilitated, and will continue to enable, global analyses of TCR recognition of pMHC, which in turn will further enhance our understanding of the drivers of MHC restriction of TCRs.

Evidence from the pre-selection repertoire

The extent to which thymic selection or germline-encoded motifs drive TCR recognition of MHC molecules can be inferred by analysis of the pre-selection TCR repertoire. Of the naturally generated TCRs in a mouse, 15–30% are activated by pMHC molecules expressed by stimulator cells from an inbred mouse strain^{33–36}. This provides a lower-bound estimate of the physiological reactivity of pre-selection TCRs because F1 hybrid stimulator cells activate more pre-selection thymocytes than do stimulator cells from inbred mice³⁴. TCRs that can be activated by more than one MHC molecule are present in the pre-selection repertoire^{33,36} and are enriched in the mature T cell repertoire of mice in which all MHC class II molecules present a single peptide^{37,38}. The extensive TCR cross reactivity for MHC that is observed in the mice expressing single-peptide–MHC class II molecules is due to a defect in negative selection in the thymus, which normally eliminates cross-reactive TCRs,

Register

The position of a peptide within the binding groove of the MHC molecule.

MHC allomorphs

Different forms of an MHC protein encoded by different MHC alleles.

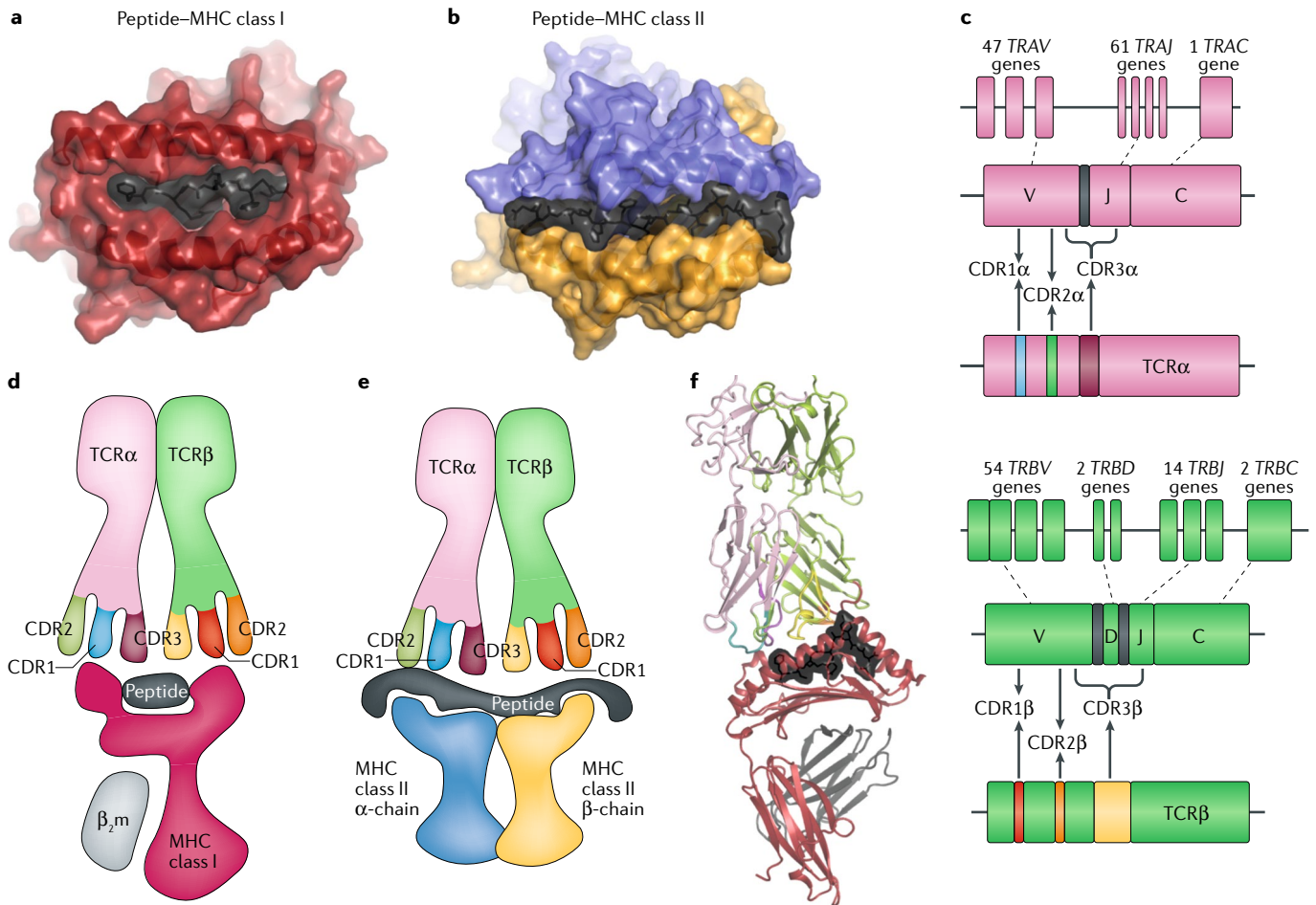


Fig. 2 | Overview of TCR recognition of peptide–MHC class I and peptide–MHC class II. **a** | Overview of peptide (black stick and surface) in complex with an MHC class I molecule (red surface). **b** | Overview of peptide (black stick and surface) in complex with an MHC class II molecule (MHC β -chain in orange and α -chain in blue). **c** | Genomic organization and recombination of T cell receptor (TCR) α -chain genes (pink) and TCR β -chain genes (green). The complementarity-determining regions (CDRs) are shown in blue, green and maroon for CDR1, CDR2 and CDR3 of the α -chain, respectively, and in red, orange and yellow for CDR1, CDR2 and CDR3 of the β -chain, respectively. CDR1 and CDR2 are encoded by TRBAV and TRABV gene segments, and CDR3 encompasses the junction between V and J regions (for the TCR α -chain) or between VD and J regions (for the TCR β -chain). Non-templated nucleotide insertions and deletions are represented by the black box. **d** | Schematic depicting typical interactions between TCR CDR loops and peptide–MHC (pMHC) class I. **e** | Schematic depicting typical interactions between TCR CDR loops and pMHC class II. In both cases, the CDR loops are coloured as per part **c**. **f** | Crystal structure of a TCR (coloured as per part **c**) in complex with pMHC class I (coloured as per part **a**) derived by using the CF34 TCR in complex with HLA-B8–FLRGRAYGL (a peptide derived from an Epstein–Barr virus immunodominant latent antigen) as a model⁷⁴. β_2m , β_2 -microglobulin.

as a result of limited ligand availability^{33,39,40}. The indication from these studies that a substantial proportion (15–30%) of the pre-selection TCR repertoire is activated by pMHC is consistent with a germline-encoded model for TCR–pMHC recognition.

Dissecting the inherent recognition capacity of TCRs for MHC molecules is complicated by the fact that successful TCR signalling requires the tyrosine-protein kinase LCK, which is typically found associated with the CD4 and CD8 co-receptors^{41,42}. During TCR–pMHC interactions, the ectodomains of CD4 or CD8 bind to the MHC molecule, while their endodomains localize LCK to its target, the CD3 signalling complex^{43,44}. In pre-selection thymocytes, only 1.80% of CD4 molecules and 0.16% of CD8 molecules carry active LCK⁴⁵, which is likely to be a key constraint on

TCR signalling. To distinguish the intrinsic capacity of the TCR to recognize MHC from its requirement for LCK, studies have uncoupled LCK from CD4 and CD8 to enable co-receptor-independent TCR signalling. This was initially achieved using mice deficient in MHC class I and class II molecules, CD4 and CD8 (known as quad-knockout mice)¹⁵ and later using mice expressing a targeted mutation of LCK¹⁶. In these mice, T cells developed that expressed non-MHC-restricted TCRs capable of recognizing conformational epitopes in a manner akin to antibody–antigen recognition¹⁴. These findings provided evidence that the MHC restriction of TCRs is, at least partly, imposed extrinsically, which supports the selection theory of TCR recognition. However, these non-MHC-restricted TCRs were disproportionately (40%) made up of

CD155-reactive TCRs, which potentially suggests that the germline-encoded recognition of MHC molecules by the TCR has been redirected towards a limited number of other antigens in these mice. Nevertheless, these CD155-reactive TCRs were not cross reactive with MHC and they recognized distinct epitopes of CD155, a molecule that is ubiquitously expressed in the thymus.

Other investigations to determine whether germline-encoded features promote TCR–pMHC interactions have involved the mutation of conserved amino acid residues in the TCR or MHC^{10,11,46}. Individual mutations of the CDR2 β residues Y46, Y48 or E54 in the TCR V β 8.2 chain, or of Y46 in the TCR V β 6 chain, markedly diminished the production of naive T cells in the thymus¹⁰, which indicates that these residues are important for the development of a T cell population of normal size. Another approach co-opted the TCR recombination machinery to randomize the CDR1 and/or CDR2 loops of the TCR α or TCR β chain. This showed that a wide range of CDR1 and CDR2 sequences and lengths could support T cell development but noted decreased production of naive T cells in the thymus, decreased expression of the TCR activation marker CD5 on pre-selection thymocytes and slower rejection of skin allografts, which are consistent with an important role for germline-encoded features in T cell selection and function⁴⁷. Conversely, mutation of outward-facing residues in the MHC class II molecule I–A^b, which were shown to be conserved TCR docking sites, had little or no effect on the number of CD4⁺ T cells and no effect on TCR diversity⁴⁸. Thus, T cell development seems to be more resilient to mutations of conserved features within the MHC than within the TCR. What has remained elusive, however, is the demonstration of generic germline-encoded motifs in the TCR and the MHC that confer recognition.

Ultimately, studies analysing the pre-selection TCR repertoire have not provided a clear answer to the question of what drives TCR recognition of MHC. Collectively, they support both the germline-encoded and selection theories of TCR recognition.

Evidence from TCR–pMHC structural studies

In 1971, Jerne postulated his views on antigen receptor diversification¹². Although many of these theories have since been shown to be incorrect^{4,5}, a central tenet of Jerne's hypothesis was an inherent evolutionary bias of the germline-encoded regions of TCRs towards recognizing MHC molecules. If we consider this in a structural context, it implies that the V gene-encoded regions of the TCR are 'hard-wired' to interact with the MHC and that the CDR3 loops 'readout' the peptide cargo⁸. By contrast, the more recently proposed selection model contests that germline-encoded TCR recognition motifs for MHC molecules are not required, as the process of thymic selection of TCRs provides the MHC reactivity¹⁷. Here, we highlight in a chronological manner key findings from TCR–pMHC structural studies (FIG. 1; Supplementary Table 1) and discuss them in the context of the two models of MHC restriction.

The 1990s. In 1996, the first structures of TCR–pMHC class I complexes were reported^{49,50}. These pioneering studies set the scene for establishing testable hypotheses about TCR recognition of pMHC. A consensus view formed whereby the TCR bound pMHC in a diagonal docking mode, with the TCR α -chain and β -chain being positioned over the α 2-helix and α 1-helix of the MHC class I molecule, respectively (FIG. 2d–f). This binding mode enabled the germline-encoded regions of the TCR to contact the MHC molecule, with the hypervariable CDR3 loops of the TCR being centred over the peptide. These structures provided immediate insight into the co-recognition of peptide and MHC molecule by the TCR. Although some deviations from this dichotomous role for the CDR1 and CDR2 loops and the CDR3 loops were noted in these initial reports⁴⁹, these structures reinforced the concept of germline-encoded TCR–MHC recognition and directly aligned with Jerne's hypothesis. A subsequent study provided important insight into TCR cross reactivity towards different peptides, whereby a conformational change of the CDR3 loop was observed upon pMHC engagement⁵¹. Soon after, additional studies showed how TCRs with different TCR gene usage engaged the same pMHC molecule and maintained the diagonal docking mode⁵², and how the germline-encoded regions of the TCR were the key energetic determinants of the interaction⁵³.

Just before the turn of the century, the first structure of a TCR–pMHC class II complex was reported. This demonstrated an orthogonal docking mode of the TCR over the MHC, which suggested that there are fundamental differences in docking geometries between TCRs from CD4⁺ and CD8⁺ T cells that are influenced by their co-receptors, CD4 and CD8, respectively⁵⁴. Collectively, these early structural studies, while demonstrating the variability of the interaction, supported the germline-encoded model of TCR–pMHC recognition.

2000–2010. In the first decade of the twenty-first century, many distinct TCR–pMHC structures were reported, which addressed key concepts of TCR cross reactivity^{55,56}, TCR bias^{57,58}, the effects of HLA polymorphism⁵⁹, TCR recognition of tumour antigens^{60,61}, alloreactivity^{62,63} and autoreactivity^{64,65}. Flexibility of the TCR CDR3 loop was shown to contribute to degeneracy of pMHC recognition, and subsequent studies showed that the CDR1 and CDR2 loops, as well as the pMHC complex itself⁶⁶, can undergo conformational change upon TCR–pMHC ligation⁵⁷. However, the CDR loops of some TCRs are relatively rigid upon binding to certain pMHC structures, which indicates that CDR loop plasticity is not necessarily a general feature of TCR–pMHC recognition⁶⁷.

The first insight into how TCRs can recognize long MHC class I-restricted peptides was provided by a structure showing a peptide-centric TCR interaction that made limited contacts with the MHC molecule itself⁶⁸. Analysis of the TCR–pMHC database at that time indicated that three MHC class I positions (65, 69 and 155; and equivalent positions in MHC class II) were invariably contacted by the TCR, which

TCR bias

Preferential usage of T cell receptors (TCRs) with specific characteristics, including gene segment usage and/or complementarity-determining region 3 (CDR3) sequence, that is typically observed in antigen-specific TCR repertoires.

Degeneracy

The ability of a T cell receptor to recognize more than one peptide–MHC complex.

suggested that these were the minimal requirements of MHC restriction. However, subsequent mutational and structural studies showed that this restriction triad was dispensable and accordingly did not represent a cardinal feature of pMHC recognition nor evidence of a germline-encoded MHC motif that directs TCR recognition⁶⁹.

Structural studies shed light on how TCRs recognize featureless peptides bound within the MHC^{58,70,71}. TCRs targeting such peptides often exhibited reproducible patterns of TCR gene segment bias, which was intriguing in the context of the germline-encoded model, as conserved *TRAV* and/or *TRBV* usage might have predicted preferred contacts with the MHC molecule. Although some of this TCR bias could be attributed to MHC contacts, such germline-encoded regions of the TCR were frequently observed to contact the peptide or were attributed to preferential TCR chain pairing⁷². Furthermore, it was established through mutagenesis studies that the CDR3 loops of the TCR could be the energetic drivers of the interaction with the MHC molecule and/or peptide⁷³. Collectively, these studies provided evidence against an inherent bias of the germline TCR sequence to recognize MHC molecules by showing that the germline-encoded regions of the TCR can be predisposed towards binding to the peptide itself, with a wide range of docking geometries underpinning such recognition.

The suggestion from earlier studies of a generic difference between the docking geometries of TCR–MHC class I and TCR–MHC class II complexes (diagonal and orthogonal, respectively) was subsequently proved to be incorrect. An autoreactive TCR–MHC class II complex revealed extreme amino-terminal positioning of the TCR over the antigen-binding platform of the MHC molecule, which suggested a link between atypical TCR docking modes and autoreactivity⁶⁴. However, other autoreactive TCR–pMHC complexes adopted more standard docking modes, and antimicrobial TCR ternary complexes could also have atypical docking modes⁷⁴.

Insights into T cell alloreactivity were also gained. Historically, two theories had been considered, namely, peptide-centric alloreactivity and MHC-centric alloreactivity. One study in this period showed that alloreactivity could be attributed to the TCR adopting two distinct docking modes over the pMHC⁶², whereas another study supported peptide-centric alloreactivity⁶³. Therefore, the inherent variability of TCR–pMHC recognition provided evidence in support of both theories. Collectively, these studies showed that there is a large degree of variability in TCR recognition of pMHC and in doing so invalidated some early models that had aligned the nature of pMHC recognition with specific functional outcomes for T cells.

Despite the substantial variation in TCR–pMHC recognition that was revealed by the structural studies, two previously held generalizations remained: namely, the need for the TCR to co-recognize the peptide and the MHC molecule and the consensus polarity of the TCR atop the MHC, which was a key tenet of Jerne's original hypothesis⁵. Within this conceptual framework, a series of investigations involving V β 8.2⁺ TCRs, and

both MHC class I and MHC class II molecules, documented conserved pairwise interactions, ostensibly between the CDR2 β loop and the MHC molecules^{7,8,10,75}. These interactions, which were found to be largely conserved across species, were taken as strong evidence for the germline-encoded regions of the TCR having inherent MHC reactivity. Nevertheless, it was observed that V β 8.2⁺ TCRs could interact with different regions of the MHC, and these variations were attributed to differing TCR sequences, differing MHC allotypes or differing peptides presented by the same MHC molecule (FIG. 3a). Moreover, mice were recently generated in which several key residues of the MHC class II molecule I–A^b, which mediate interactions with these conserved TCR motifs, were mutated to abrogate TCR binding. T cells in these mice developed normally and generated large diverse repertoires, albeit with altered *TRAV* and *TRBV* usage relative to wild-type mice⁴⁸. These data suggest, at the least, that there is a lack of universality of such germline-encoded TCR motifs. As an alternative to preferred pairwise interactions between TCRs and MHC molecules, it has recently been suggested that biophysical parameters, including charge or shape complementarity between the TCR and MHC molecule, can function as conserved molecular drivers of this interaction⁷⁶. The conserved TCR docking polarity, coupled with the existence of conserved motifs (albeit less than universal), was supportive of an inherent bias of TCRs towards recognizing MHC molecules.

2010 to date. So far, 53 unique TCR–pMHC class I complexes and 26 unique TCR–pMHC class II complexes have been determined (Supplementary Table 1) from a total of 172 TCR–pMHC structures currently deposited in the [Protein Data Bank \(PDB\)](#)⁷⁷. Studies in the past decade have provided key insight into biased TCR usage in the context of protective immunity and aberrant reactivity^{71,78–83}, more examples of autoreactive TCR ternary complexes^{84–87}, examples of MHC polymorphism shaping TCR recognition^{88,89}, insight into how TCR cross reactivity towards differing peptides can be attributed to highly focused molecular mimicry (as observed in peptide-display library approaches and autoreactive disease settings^{90,91}) and the first example of a TCR having cross reactivity for MHC class I and MHC class II molecules⁸⁶. This last study simultaneously highlighted the adaptability of the TCR and strengthened the concept of preferred interaction codons as demonstrated by the observed consensus polarity of TCR docking atop the MHC molecules⁸⁶. The codon concept was expanded to suggest that murine V α 3.3⁺ TCRs are predisposed to interact with a defined region of H2–L^d (REF.⁹²). Indeed, deviations from the TCR–pMHC docking geometry as determined by the interaction codon correlated with poor signalling, thereby providing a functional link between preferred germline-encoded TCR–MHC contacts and TCR signalling⁹³. However, the variation of docking geometry observed in this study fell well within the observed overall range of TCR–pMHC docking geometries, which suggests that additional factors contributed to the poorer signalling outcome. Moreover, other studies showed that identical TCR β -chains can have different pMHC binding modes⁹⁴, and the V α 3.3⁺

Ternary complexes

Protein complexes containing three different molecules bound together — namely, the T cell receptor, peptide and an MHC molecule.

Pairwise interactions

Conserved interactions between particular residues on the MHC molecule with paired or matching residues on the T cell receptor.

Molecular mimicry

Similarity in peptide sequences that is sufficient to induce cross reactivity among T cell receptors.

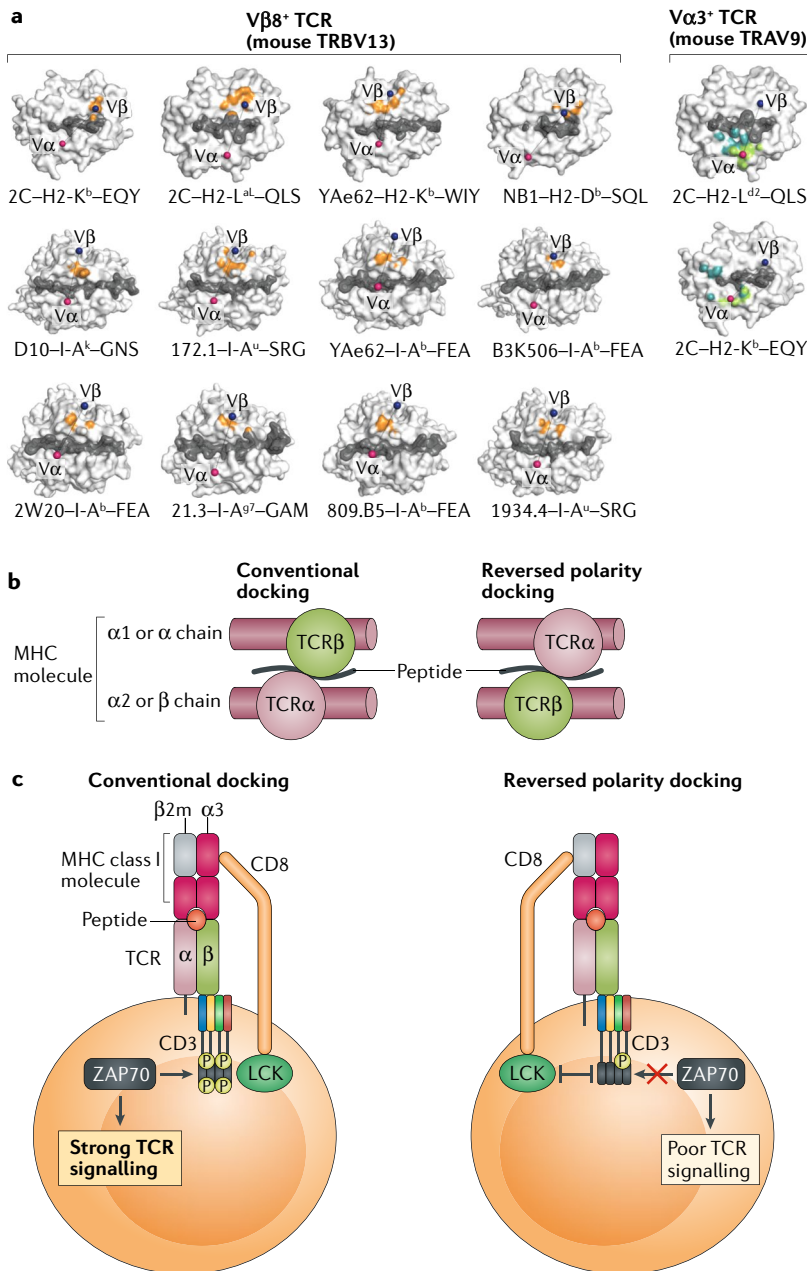


Fig. 3 | Conventional and reversed polarity TCR docking on MHC molecules. **a** | V β 8⁺ and V α 3⁺ T cell receptor (TCR) recognition. Surface representation of MHC class I and MHC class II molecules (white surface) with their bound peptide (grey surface) solved in complex with mouse V β 8⁺ TCRs (equivalent to TCR β -chain variable 13 (TRBV13) TCRs) or mouse V α 3⁺ TCRs (equivalent to TCR α -chain variable 9 (TRAV9) TCRs). Each structure is labelled according to the convention of TCR clone–MHC type–peptide (first three amino acids in peptide provided here; full sequences provided in Supplementary Table 1). The surface area that interacts with the TCR β complementarity-determining region 2 (CDR2 β) loop is shown in orange, with the CDR2 α loop is shown in green and with the CDR1 α loop is shown in teal. The pink and blue spheres represent the mass centre of the V α and V β domains, respectively. These footprints show the preferred pairwise interactions between MHC molecules and TCRs that have been reported and how these can vary among individual peptide–MHC (pMHC) complexes. **b** | Schematic representation of conventional TCR docking and reversed polarity TCR docking. **c** | Model depicting how conventional TCR–pMHC class I docking topology is proposed to facilitate the appropriate apposition of key signalling molecules and substrates, namely, the tyrosine–protein kinase LCK and CD3 immunoreceptor tyrosine-based activation motifs, and thus support effective TCR-mediated signalling¹⁰³. Reversed polarity TCR–pMHC class I docking^{97,98} prevents this association, resulting in diminished or abrogated signalling. β 2m, β 2-microglobulin.

TCRs have disparate recognition modes with other MHC molecules (FIG. 3a).

Collectively, it has been shown that various TCR docking modalities are associated with pMHC recognition, with many examples of CDR3–peptide and CDR3–MHC interactions and of how the CDR3 loops can alter TCR–MHC contacts despite conserved TCR gene usage. Furthermore, it was shown that TRBV allelic polymorphism directly affects peptide reactivity⁸³. Consistent with this, in two recent studies, the TRBV chain directly contacted the peptide, which determined the observed TCR bias and functional outcome^{95,96}. Nevertheless, the need for co-recognition of peptide and MHC molecule, together with the common docking polarity of the TCR over the MHC molecule (FIG. 3b), remains a generality of TCR–pMHC recognition, thereby supporting Jerne’s key hypothesis of an evolutionary bias of the germline-encoded regions of the TCR towards MHC binding.

Thus, it was of major importance that two distinct examples of reversed TCR docking topologies, for MHC class II and, later, MHC class I molecules, were reported^{97,98}. First, two TCRs isolated from human peripherally derived regulatory T (pT_{reg}) cells were shown to bind to MHC class II molecules in a 180° reversed orientation, with the TCR α -chain and β -chain positioned over the α -chain and β -chain of the MHC class II molecule, respectively⁹⁷ (FIG. 3b). Here, the TCR α -chain did not contact the pMHC, and the germline-encoded regions of the TCR β -chain solely contacted the self-peptide. Notably, TCRs having the same V gene segments as this pT_{reg} cell TCR were shown to adopt consensus TCR–pMHC docking topologies. Second, a reversed TCR–pMHC class I docking polarity was observed for two TCRs identified from the pre-immune TCR repertoire in mice⁹⁸. These TRBV17⁺ TCRs were also 180° reverse-orientated with respect to the MHC molecule, whereby the TCR α -chain and β -chain were located over the α 1 and α 2 helix of the MHC class I molecule, respectively. Notably, the peptide solely interacted with the germline-encoded TRBV17 region, and no contact was observed between any of the TCR CDR3 loops and the peptide. Despite these unusual binding characteristics, the pre-immune TCR bound the pMHC class I molecule with reasonable affinity and elicited a TCR signal, although the reversed docking mode was correlated with relatively poor signalling and a minimal expansion of the TRBV17⁺ T cell population in the immune repertoire.

These two examples of reversed TCR docking require a reappraisal of what docking polarity tells us about MHC restriction and germline-encoded recognition. Before these observations, the absence of a reversed docking TCR–pMHC polarity was put forward as strong evidence in favour of the germline-encoded model^{8,9,99}; in other words, the reproducible manner in which TCRs were observed to interact with MHC molecules strongly indicated the presence of evolutionarily conserved contacts. Presently, of the approximately 80 unique TCR–pMHC structures that have been determined (Supplementary Table 1), two have exhibited a reversed polarity. Looking merely at the relative frequencies of conventional and reversed TCR docking structures, it is tempting to class these examples as

outliers to the general view of TCRs being biased towards MHC recognition. However, it is necessary to appreciate the context in which the reversed TCRs were observed. The vast majority of the TCR–pMHC structural information has focused on TCRs that have originated from the immune repertoire, with a heavy emphasis on certain MHC molecules in this structural database (for example, HLA-A2 represents 35% of the structures in the database) (Supplementary Table 1), and thus could be argued to represent deep sampling of a narrow pool. The reversed polarity MHC class I-restricted TCR was the first ternary structure of an antigen-specific TCR from an unexpanded (naive) repertoire, which suggests that the frequency of unconventionally docking TCRs is under-represented in the structural database and highlights the need to sample the TCR repertoire more broadly. Moreover, on the basis of the reversed polarity TCR–MHC class II complex from the pT_{reg} cell, the field needs to resolve more T_{reg} cell TCR ternary complexes to establish whether reversed docking is a common feature underpinning T_{reg} cell biology.

Although they have been identified in endogenous repertoires, the reversed docking TCRs signal poorly, which is likely to explain their poor representation in immune cell populations. Thus, although unconventional docking is possible, it may not be an optimal recognition modality for signal transduction (discussed below). Nevertheless, one generality remains in TCR–pMHC recognition, namely, the obligate need for the TCR to simultaneously recognize the MHC molecule and the peptide.

Evidence from TCR signalling studies

Productive TCR co-recognition of pMHC depends on downstream signalling molecules that are activated by this recognition event. Below, we discuss the mechanisms by which it is suggested that MHC restriction, and in particular, the conserved positioning of the TCR over the MHC molecule, is driven by signalling constraints that are imposed by the need for key signalling molecules to interact with their substrates. It is these signalling constraints that underpin the selection model, and much of the evidence presented in this section illuminates the importance of thymic selection in generating an MHC-focused TCR repertoire.

The selection model of TCR–pMHC recognition posits that MHC restriction is a direct consequence of the need for the TCR–CD3 complex to access LCK. As LCK is largely associated with CD4 and CD8 (especially in thymocytes), its delivery to the TCR–CD3 complex is dependent on binding of both CD4 or CD8 and TCR to the MHC molecule. Thus, MHC restriction is proposed to arise through a process that selects for TCRs that colocalize with co-receptor-bound LCK⁴⁵, with non-MHC-reactive TCRs being unable to generate a productive signal, irrespective of ligation. This theory was supported by studies (described earlier) in which LCK was liberated from the CD4 and CD8 co-receptors and could thus support TCR-mediated signal transduction independently of the nature of the ligand^{15,16}. The non-MHC-restricted TCRs that were identified in these mice indicated that the constraints around

TCR-mediated signal transduction contributed, in part, to MHC restriction. A later study tethered LCK to CD4 with the goal of augmenting TCR-mediated signals and thereby reducing the threshold for selection. Here, MHC class II-restricted TCRs gained the capacity to be activated by different peptides and MHC class II molecules, whereas MHC class I-restricted TCRs gained the capacity to be activated by MHC class II molecules¹⁰⁰. These data were interpreted as indicating the capacity of TCRs for subthreshold recognition of pMHC independently of MHC class, allele or bound peptide, which is suggestive of a TCR-intrinsic mechanism of MHC recognition. However, a caveat of this study is that the TCRs investigated were post-selection TCRs, which have a well-characterized extent of MHC binding. Thus, the TCR cross reactivity observed may be more reflective of the similarities among MHC molecules than an underlying predilection on the part of the TCR for recognition of MHC molecules.

Precisely how could the requirement for LCK determine the highly conserved docking polarity of the TCR over the pMHC? This may be related to the necessary juxtaposition of signalling molecules that is required for effective TCR-mediated signal transduction. Although the architecture of the TCR–pMHC–CD3–CD4 (or CD8) complex has not been elucidated, resolution of a TCR–pMHC class II–CD4 ternary complex showed that it has an arch-like structure that enables simultaneous engagement of TCR and CD4 by the MHC class II molecule¹⁰¹, and further studies localized the CD3 complex within the arch bound to the TCR β -chain¹⁰². This formation ensures proximity between LCK and CD3, which enables signal propagation (FIG. 3c). Reversal of the TCR–pMHC docking topology (as discussed above) would likely position CD3 outside of the arch and away from LCK, which would potentially diminish or abrogate the TCR-mediated signal¹⁰³. Although both examples of reversed TCR docking can signal, the signal intensity transduced by this interaction seems to be reduced relative to the affinity of the TCR–pMHC interaction. Interestingly, both reversed TCRs dock in a position rotated exactly 180° from the consensus polarity docking mode (FIG. 3b), which suggests that any constraints that are imposed on TCR–pMHC docking with respect to signal propagation are satisfied in either orientation. Moreover, some T cells are co-receptor independent²⁰ and several naturally occurring $\alpha\beta$ TCRs are activated by antigen completely independently of an MHC molecule^{104–108}, which makes it challenging to account for the LCK-proximity model in these T cells. Nevertheless, constraints on TCR–MHC-mediated signalling provide an explanation for how TCRs with randomly generated specificities would, following thymic selection, be exquisitely targeted towards MHC reactivity.

Evidence from systems immunology

How can global analytical approaches and recent advances in the ability to generate and interpret large data sets improve our understanding of the effective drivers of TCR recognition of MHC molecules? Potentially, germline-encoded pairwise recognition motifs in TCRs and MHC molecules would result in

TCR biases associated with the expression of particular MHC alleles. Multiple studies, including the use of high-throughput sequencing to provide global repertoire analyses, have shown that there are reproducible differences in V α and V β usage between CD4⁺ and CD8⁺ T cell subsets, which suggests that, at the least, particular V regions preferentially bind to MHC class I or MHC class II molecules^{109–112}. Associations between TCR gene usage and the expression of MHC allelic variants were not as obvious, however. Although some studies showed substantial similarities in V gene usage in HLA-identical siblings¹¹⁰, these studies focused on twins, in which similar V gene usage was observed even before thymic selection. This suggests that the TCR similarity was driven largely by shared genes involved in the TCR recombination machinery, rather than by shared MHC allelic expression^{113,114}.

Recently, two papers correlated the expression of particular TCR V genes or CDR3 β sequences with genetic variation in MHC expression in humans^{115,116}. An advantage of these studies was the large sample sizes, which enabled robust analyses of TCR–MHC associations while avoiding the complete genetic identity that confounds twin studies. One study used expression quantitative trait locus mapping to demonstrate a correlation between *TRAV* gene usage and HLA type in humans¹¹⁶. Furthermore, the TCR residues largely responsible for the correlation were clustered near the MHC contact interface and were involved in interaction with either the MHC molecule or the peptide, which indicates that the TCR–pMHC interaction underpinned this correlation. The second study, using high-throughput sequencing of TCR CDR3 β sequences from more than 600 individuals, showed a robust and predictive association between the expression of particular CDR3 β sequences and HLA type¹¹⁵. Interestingly, the demonstrated association between V gene usage and MHC allelic expression, while likely reflecting to some extent preferential interactions directly between the TCR and MHC molecules, may also correspond to a bias in TCR binding of the peptide repertoire presented by distinct MHC alleles. It is also possible that the MHC-bound peptide repertoire itself has exerted evolutionary pressure on the TCR. Some support for this concept comes from a recent study showing that the germline-encoded V gene elements are immune response genes that are required for T cell reactivity to a murine malaria epitope⁹⁶.

Characterization of the TCR repertoire is increasing at an unprecedented rate, owing in large part to the advent of high-throughput TCR sequencing (FIG. 1; Supplementary Figure 1), which has resulted in millions of TCR sequences being made available in public databases. The immediate benefit of analysing TCR sequences outside of the context of antigen specificity may seem limited with respect to understanding TCR recognition of pMHC. However, a recent network analysis¹¹⁷ of high-throughput sequencing data of TCR β from mice and humans showed that there are high levels of similarity in TCR repertoire structures of healthy individuals, in which networks of highly related CDR3 regions centred around public sequences. As a result, the TCR repertoire was more restricted

than would arise from random somatic recombination. Intriguingly, this ordered structure was found to be imparted to a large extent by thymic selection processes, with CDR3 β sequences from pre-selection double-negative thymocytes, as well as those from quad-knockout mice (mentioned earlier)¹⁵, found to be substantially less connected. This global analysis suggests, in part, that thymic selection has a key role in establishing the defining characteristics of the pre-immune TCR repertoire¹¹⁷.

The utility of data from antigen-specific, rather than total, TCRs lies in the ability to connect TCR sequences, biases and preferential chain and gene element combinations with antigen specificity. Although high-throughput sequencing is less commonly applied to antigen-specific TCRs, recent advances in the detection of paired $\alpha\beta$ TCR sequences in particular (FIG. 1; Supplementary Figure 1) have underpinned the rapid rise in available data sets. Such information can then conceivably be used to predict antigen specificity from unrelated or uncontextualized TCR information. Two recent studies have done just that using databases of multiple antigen-specific TCR sequences to develop algorithms to predict antigen specificity, with a remarkable degree of accuracy^{118,119}. Both studies relied on the generation of training data sets to predict novel TCRs that shared the same antigen specificity. These approaches exploited the fact that TCRs that bind to the same epitope share several quantifiable sequence features. Both studies worked directly from sequence data, although the choice of sequences and construction of the algorithm were informed by structural insights into the regions of the TCR that are most likely to influence pMHC recognition.

Even with extensive training sets, algorithms such as these were not able to correctly categorize all the antigen-specific TCRs that respond to a particular epitope. One of the studies identified a substantial proportion of TCR clones ('outliers') within each antigen-specific repertoire whose extreme diversity precluded their contribution to any predictive algorithm¹¹⁸. An area for future development is to investigate whether these outlier TCR sequences share 3D structural features that can be quantified and that bring them into the same 'cluster' as the more conventionally similar receptors within an epitope-specific response. Moreover, of particular relevance to the two models that have been proposed to underpin MHC restriction, it remains to be seen whether such algorithms could be refined such that the MHC restriction element could be predicted from a random assortment of TCRs independently of the bound peptide. The development of such algorithms could facilitate the identification of germline-encoded interaction motifs.

Conclusions

Understanding the extent to which evolutionary versus developmental processes shape TCR recognition of MHC molecules is more than academic. On the surface, it advances our fundamental knowledge of T cell development and the precise mechanism by which T cells are activated. At a deeper level, it provides information on the capacity of the system (the nature of TCRs that

Expression quantitative trait locus

A genetic locus that contributes to variation in expression levels of particular genes.

Public sequences

T cell receptor sequences that are often found across multiple individuals.

Proximate causation

The immediate influences on an outcome, for example, thymic selection of T cell receptors that can recognize MHC molecules.

Ultimate causation

The distal or evolutionary influences on an outcome, for example, the evolution of germline-encoded T cell receptor recognition of MHC molecules.

are possible), as well as offering targets for intervention when the system breaks down, as in autoimmunity, or opportunities for better-informed augmentation of TCR–pMHC recognition, as is the aim of current cancer immunotherapies.

The selection model of MHC restriction proposes that randomly generated TCR specificities become focused on recognizing MHC molecules during thymic selection, whereas the germline-encoded model posits that TCRs have an inherent reactivity for MHC molecules. How does one reconcile these differing explanations in the face of all the available data? Although considerable insights have been gained from experiments designed to distinguish between the two models, the results do not exclude either possibility, and sporadic TCR–pMHC structural studies over the next decade will probably not provide a clear answer.

Improvements in our capacity to predict TCR epitope specificity or MHC restriction will rely on improved structural understanding of the TCR–pMHC interaction. Crystal structures of TCRs in complex with their specific pMHC are the gold standard, providing a

definitive answer to the questions posed by TCR sequence data. However, unlike TCR sequencing, the resolution of TCR–pMHC crystal structures will probably not evolve into a high-throughput process. Accordingly, a broader sampling of TCR–pMHC complexes is required to provide an unbiased picture of the potential of TCR–pMHC recognition. The ability to segregate TCRs rationally into related clusters, as described above, provides a method for the efficient selection of TCR–pMHC complexes for structural determination that might maximize new knowledge. Although targeted selection of TCR structures will be important, the growing dichotomy between the challenge to obtain structural information and the ready ability to sequence antigen-specific TCRs highlights the need to consider alternative approaches, such as predictive algorithms, to fully interrogate the TCR repertoire in the context of understanding the structural correlates of MHC restriction.

As has been suggested previously^{13,99}, this conceptual controversy may be clarified by recognizing that the germline-encoded and selection models of MHC restriction are far from being mutually exclusive; rather, they are complementary models that, we propose, describe two causations of the same phenomenon¹²⁰ (FIG. 4). As described by Ernst Mayr more than 50 years ago¹²⁰, a proximate causation describes immediate effects driven by physiology, whereas an ultimate causation describes the effects of evolutionary pressure. Thus, the evolution of an inherent capacity for TCRs to recognize MHC molecules is an ultimate causation of MHC restriction, whereas thymic selection processes are a proximate causation. Distinguishing these two types of causation helps to clarify the scope of each model when analysing the drivers of TCR specificity for MHC molecules. The germline-encoded model describes processes that predispose pre-selection thymocytes to recognize self-peptide–MHC ligands so that enough pre-selection thymocytes mature into T cells despite the immense MHC polymorphism within a species. The selection model explains how the mature T cell pool within any given organism comes to be filled by cells expressing TCRs that recognize the particular MHC molecules expressed by that organism. Assimilating aspects of both models accounts for the fact that some TCRs exhibit MHC recognition in a manner that does not necessarily involve interaction codons, as well as the possibility of germline-encoded recognition of peptide determinants. Within this conceptual framework, there will be a spectrum of MHC reactivity and peptide centricity, the extent of which will vary from system to system. When we consider that the interaction between broadly similar TCRs and pMHC complexes has occurred for many millions of years, juxtaposed with the extraordinary diversity of all components of this interaction, it seems intuitive that an ideal system would integrate both the focusing conferred by evolutionary constraints and the flexibility associated with random gene recombination and selection. Systems-based analyses, together with more targeted structural studies, will be required to test this new model of TCR–pMHC recognition.

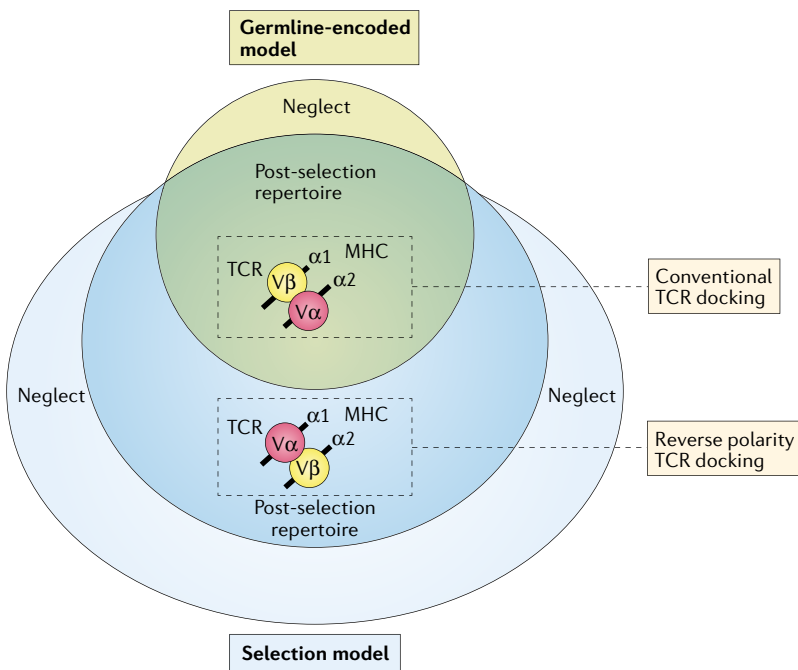


Fig. 4 | Dual causation of MHC restriction of TCRs. Currently, two models have been proposed to explain the drivers of T cell receptor (TCR) specificity for MHC molecules: the germline-encoded model and the selection model (BOX 1). Neither model has been excluded despite extensive research, so we and others^{13,99} suggest that the two models are not mutually exclusive. Rather, we propose that they describe the ‘ultimate’ and ‘proximate’ causations of TCR specificity for MHC. The schematic shows how the germline-encoded and selection models combine to produce a mature T cell repertoire with diverse TCR–MHC docking modes. The germline-encoded model describes conventional docking modes, with V α and V β loops of the TCR interacting with α 2 and α 1 helices of the MHC (class I) molecule, respectively. The selection model also allows for unconventional interactions, such as reverse polarity TCR–MHC docking, to be included in the mature T cell pool so long as productive TCR signalling is maintained in the T cell. Whereas a key distinction between these models is the nature of the pre-selection repertoires, resulting in a larger (and distinct) pool of neglected thymocytes in the selection model than in the germline model, an important feature of this framework is that an individual TCR can conform fully to both germline-encoded and selection models.

Published online: 10 April 2018

1. Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702 (1974).
2. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308**, 149–153 (1984).
3. Yanagi, Y. et al. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* **308**, 145–149 (1984).
4. Rossjohn, J. et al. T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169–200 (2015).
5. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
6. van der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* **11**, 47–55 (2011).
7. Feng, D., Bond, C. J., Ely, L. K., Maynard, J. & Garcia, K. C. Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nat. Immunol.* **8**, 975–983 (2007).
This study provides the first structural evidence of the interaction codon.
8. Garcia, K. C., Adams, J. J., Feng, D. & Ely, L. K. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat. Immunol.* **10**, 143–147 (2009).
9. Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L. & Kappler, J. W. Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
10. Scott-Browne, J. P., White, J., Kappler, J. W., Gapin, L. & Marrack, P. Germline-encoded amino acids in the alpha beta T-cell receptor control thymic selection. *Nature* **458**, 1043–1046 (2009).
11. Yin, L., Scott-Browne, J., Kappler, J. W., Gapin, L. & Marrack, P. T cells and their eons-old obsession with MHC. *Immunol. Rev.* **250**, 49–60 (2012).
12. Jerne, N. K. The somatic generation of immune recognition. *Eur. J. Immunol.* **1**, 1–9 (1971).
13. Rangarajan, S. & Mariuzza, R. A. T cell receptor bias for MHC: co-evolution or co-receptors? *Cell. Mol. Life Sci.* **71**, 3059–3068 (2014).
14. Tikhonova, A. N. et al. alpha beta T cell receptors that do not undergo major histocompatibility complex-specific thymic selection possess antibody-like recognition specificities. *Immunity* **36**, 79–91 (2012).
15. Van Laethem, F. et al. Deletion of CD4 and CD8 coreceptors permits generation of alpha beta T cells that recognize antigens independently of the MHC. *Immunity* **27**, 735–750 (2007).
This study provides evidence for the selection theory of TCR recognition.
16. Van Laethem, F. et al. Lck availability during thymic selection determines the recognition specificity of the T cell repertoire. *Cell* **154**, 1326–1341 (2013).
17. Van Laethem, F., Tikhonova, A. N. & Singer, A. MHC restriction is imposed on a diverse T cell receptor repertoire by CD4 and CD8 co-receptors during thymic selection. *Trends Immunol.* **33**, 437–441 (2012).
18. Yewdell, J. W. & Haeryfar, S. M. Understanding presentation of viral antigens to CD8⁺ T cells *in vivo*: the key to rational vaccine design. *Annu. Rev. Immunol.* **23**, 651–682 (2005).
19. Petersen, J., Purcell, A. & Rossjohn, J. Post-translationally modified T cell epitopes: immune recognition and immunotherapy. *J. Mol. Med.* **87**, 1045–1051 (2009).
20. Godfrey, D. I., Uldrich, A. P., McCluskey, J., Rossjohn, J. & Moody, D. B. The burgeoning family of unconventional T cells. *Nat. Immunol.* **16**, 1114–1123 (2015).
21. Van Rhijn, I., Godfrey, D. I., Rossjohn, J. & Moody, D. B. Lipid and small-molecule display by CD1 and MR1. *Nat. Rev. Immunol.* **15**, 643–654 (2015).
22. Rossjohn, J., Pellicci, D. G., Patel, O., Gapin, L. & Godfrey, D. I. Recognition of CD1d-restricted antigens by natural killer T cells. *Nat. Rev. Immunol.* **12**, 845–857 (2012).
23. Bjorkman, P. J. et al. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
24. Burrows, S. R., Rossjohn, J. & McCluskey, J. Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol.* **27**, 11–16 (2006).
25. Brown, J. et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39 (1993).
26. Adams, E. J. & Luoma, A. M. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. *Annu. Rev. Immunol.* **31**, 529–561 (2013).
27. Hendersen, K. N. et al. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* **27**, 23–34 (2007).
28. Smith, K. J. et al. An altered position of the alpha 2 helix of MHC class I is revealed by the crystal structure of HLA-B*3501. *Immunity* **4**, 203–213 (1996).
29. Tynan, F. E. et al. High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T cell receptor engagement and T cell immunodominance. *J. Biol. Chem.* **280**, 23900–23909 (2005).
30. Turner, S. J., Doherty, P. C., McCluskey, J. & Rossjohn, J. Structural determinants of T-cell receptor bias in immunity. *Nat. Rev. Immunol.* **6**, 883–894 (2006).
31. Lefranc, M. P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **29**, 207–209 (2001).
32. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
33. McDonald, B. D., Bunker, J. J., Erickson, S. A., Oh-Hora, M. & Bendelac, A. Crossreactive alpha beta T cell receptors are the predominant targets of thymocyte negative selection. *Immunity* **43**, 859–869 (2015).
34. Merckenschlager, M. et al. How many thymocytes audition for selection? *J. Exp. Med.* **186**, 1149–1158 (1997).
35. Sinclair, C., Bains, I., Yates, A. J. & Seddon, B. Asymmetric thymocyte death underlies the CD4:CD8 T-cell ratio in the adaptive immune system. *Proc. Natl Acad. Sci. USA* **110**, E2905–E2914 (2013).
36. Zerrahn, J., Held, W. & Raulat, D. H. The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636 (1997).
37. Huseby, E. S. et al. How the T cell repertoire becomes peptide and MHC specific. *Cell* **122**, 247–260 (2005).
38. Ignatowicz, L., Kappler, J. & Marrack, P. The repertoire of T cells shaped by a single MHC/peptide ligand. *Cell* **84**, 521–529 (1996).
39. Chu, H. H., Moon, J. J., Kruse, A. C., Pepper, M. & Jenkins, M. K. Negative selection and peptide chemistry determine the size of naive foreign peptide-MHC class II-specific CD4⁺ T cell populations. *J. Immunol.* **185**, 4705–4713 (2010).
40. Huseby, E. S., Crawford, F., White, J., Kappler, J. & Marrack, P. Negative selection imparts peptide specificity to the mature T cell repertoire. *Proc. Natl Acad. Sci. USA* **100**, 11565–11570 (2003).
41. Turner, J. M. et al. Interaction of the unique N-terminal region of tyrosine kinase p56lck with cytoplasmic domains of CD4 and CD8 is mediated by cysteine motifs. *Cell* **60**, 755–765 (1990).
42. Veillet, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck. *Cell* **55**, 301–308 (1988).
43. Artyomov, M. N., Lis, M., Devadas, S., Davis, M. M. & Chakraborty, A. K. CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery. *Proc. Natl Acad. Sci. USA* **107**, 16916–16921 (2010).
44. Li, Q. J. et al. CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. *Nat. Immunol.* **5**, 791–799 (2004).
45. Stepanek, O. et al. Coreceptor scanning by the T cell receptor provides a mechanism for T cell tolerance. *Cell* **159**, 333–345 (2014).
46. Scott-Browne, J. P. et al. Evolutionarily conserved features contribute to ab t cell receptor specificity. *Immunity* **35**, 526–535 (2011).
47. Holland, S. J. et al. The T-cell receptor is not hardwired to engage MHC ligands. *Proc. Natl Acad. Sci. USA* **109**, E3111–E3118 (2012).
48. Silberman, D. et al. Class II major histocompatibility complex mutant mice to study the germ-line bias of T-cell antigen receptors. *Proc. Natl Acad. Sci. USA* **113**, E5608–E5617 (2016).
49. Garboczi, D. N. et al. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134–141 (1996).
50. Garcia, K. C. et al. An alpha beta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**, 209–219 (1996).
References 49 and 50 provide the first molecular snapshots of the TCR–pMHC interaction.
51. Garcia, K. C. et al. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* **279**, 1166–1172 (1998).
52. Ding, Y. H. et al. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* **8**, 403–411 (1998).
53. Manning, T. C. et al. Alanine scanning mutagenesis of an alpha beta T cell receptor: mapping the energy of antigen recognition. *Immunity* **8**, 413–425 (1998).
54. Reinherz, E. L. et al. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* **286**, 1913–1921 (1999).
55. Reiser, J. B. et al. A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity* **16**, 345–354 (2002).
56. Reiser, J. B. et al. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat. Immunol.* **4**, 241–247 (2003).
57. Kjer-Nielsen, L. et al. A structural basis for the selection of dominant alpha beta T cell receptors in antiviral immunity. *Immunity* **18**, 53–64 (2003).
58. Stewart-Jones, G. B., McMichael, A. J., Bell, J. I., Stuart, D. I. & Jones, E. Y. A structural basis for immunodominant human T cell receptor recognition. *Nat. Immunol.* **4**, 657–663 (2003).
59. Archbold, J. K. et al. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. *J. Exp. Med.* **206**, 209–219 (2009).
60. Deng, L. et al. Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nat. Immunol.* **8**, 398–408 (2007).
61. Chen, J. L. et al. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J. Exp. Med.* **201**, 1243–1255 (2005).
62. Colf, L. A. et al. How a single T cell receptor recognizes both self and foreign MHC. *Cell* **129**, 135–146 (2007).
63. Macdonald, W. A. et al. T cell all recognition via molecular mimicry. *Immunity* **31**, 897–908 (2009).
References 62 and 63 describe the molecular basis of T cell alleloreactivity.
64. Hahn, M., Nicholson, M. J., Pyrdol, J. & Wucherpfennig, K. W. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nat. Immunol.* **6**, 490–496 (2005).
This study provides the first insight into an autoreactive TCR–pMHC interaction.
65. Li, Y. et al. Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. *EMBO J.* **24**, 2968–2979 (2005).
66. Borbulevich, O. Y. et al. T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity* **31**, 885–896 (2009).
67. Tynan, F. E. et al. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nat. Immunol.* **8**, 268–276 (2007).
68. Tynan, F. E. et al. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nat. Immunol.* **6**, 1114–1122 (2005).
69. Burrows, S. R. et al. Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proc. Natl Acad. Sci. USA* **107**, 10608–10613 (2010).
70. Turner, S. J. et al. Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8⁺ T cell populations. *Nat. Immunol.* **6**, 382–389 (2005).
71. Day, E. B. et al. Structural basis for enabling T-cell receptor diversity within biased virus-specific CD8⁺ T cell responses. *Proc. Natl Acad. Sci. USA* **108**, 9536–9541 (2011).
72. Gras, S., Kjer-Nielsen, L., Burrows, S. R., McCluskey, J. & Rossjohn, J. T cell receptor bias and immunity. *Curr. Opin. Immunol.* **20**, 119–125 (2008).
73. Borg, N. A. et al. The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. *Nat. Immunol.* **6**, 171–180 (2005).
74. Gras, S. et al. The shaping of T cell receptor recognition by self-tolerance. *Immunity* **30**, 193–203 (2009).
75. Dai, S. et al. Crossreactive T cells spotlight the germline rules for [alpha][beta] T cell-receptor interactions with MHC molecules. *Immunity* **28**, 324–334 (2008).
76. Blevins, S. J. et al. How structural adaptability exists alongside HLA-A2 bias in the human alpha beta TCR

- repertoire. *Proc. Natl Acad. Sci. USA* **113**, E1276–E1285 (2016).
77. Berman, H. M. et al. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
78. Ladell, K. et al. A molecular basis for the control of preimmune escape variants by HIV-specific CD8⁺ T cells. *Immunity* **38**, 425–436 (2013).
79. Petersen, J. et al. Determinants of gliadin-specific T cell selection in celiac disease. *J. Immunol.* **194**, 6112–6122 (2015).
80. Broughton, S. E. et al. Biased T cell receptor usage directed against human leukocyte antigen DO8-restricted gliadin peptides is associated with celiac disease. *Immunity* **37**, 611–621 (2012).
81. Petersen, J. et al. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. *Nat. Struct. Mol. Biol.* **21**, 480–488 (2014).
82. Gras, S. et al. A structural basis for varied α TCR usage against an immunodominant EBV antigen restricted to a HLA-B8 molecule. *J. Immunol.* **188**, 311–321 (2012).
83. Gras, S. et al. Allelic polymorphism in the T cell receptor and its impact on immune responses. *J. Exp. Med.* **207**, 1555–1567 (2010).
84. Sethi, D. K. et al. A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. *J. Exp. Med.* **208**, 91–102 (2011).
85. Sethi, D. K., Gordo, S., Schubert, D. A. & Wucherpfennig, K. W. Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop. *Nat. Commun.* **4**, 2623 (2013).
86. Yin, L. et al. A single T cell receptor bound to major histocompatibility complex class I and class II glycoproteins reveals switchable TCR conformers. *Immunity* **35**, 23–33 (2011). **This paper describes how a TCR can bind both MHC class I and MHC class II molecules.**
87. Bulek, A. M. et al. Structural basis for the killing of human beta cells by CD8⁺ T cells in type 1 diabetes. *Nat. Immunol.* **13**, 283–289 (2012).
88. Stewart-Jones, G. B. et al. Structural features underlying T-cell receptor sensitivity to concealed MHC class I micropolymorphisms. *Proc. Natl Acad. Sci. USA* **109**, E3483–E3492 (2012).
89. Liu, Y. C. et al. A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism. *J. Biol. Chem.* **289**, 16688–16698 (2014).
90. Birnbaum, M. E. et al. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
91. Cole, D. K. et al. Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *J. Clin. Invest.* **126**, 3626–3626 (2016).
92. Adams, J. J. et al. Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat. Immunol.* **17**, 87–94 (2015).
93. Adams, J. J. et al. T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* **35**, 681–693 (2011).
94. Stadinski, B. D. et al. A role for differential variable gene pairing in creating T cell receptors specific for unique major histocompatibility ligands. *Immunity* **35**, 694–704 (2011).
95. Culshaw, A. et al. Germline bias dictates cross-serotype reactivity in a common dengue-virus-specific CD8⁺ T cell response. *Nat. Immunol.* **18**, 1228–1237 (2017).
96. Van Braeckel-Budimir, N. et al. A T cell receptor locus harbors a malaria-specific immune response gene. *Immunity* **47**, 835–847 (2017).
97. Beringer, D. X. et al. T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nat. Immunol.* **16**, 1153–1161 (2015).
98. Gras, S. et al. Reversed T cell receptor docking on a major histocompatibility class I complex limits involvement in the immune response. *Immunity* **45**, 749–760 (2016). **References 97 and 98 highlight the existence of reversed TCR–pMHC docking topologies.**
99. Garcia, K. C. Reconciling views on T cell receptor germline bias for MHC. *Trends Immunol.* **33**, 429–436 (2012).
100. Parrish, H. L., Deshpande, N. R., Vasic, J. & Kuhns, M. S. Functional evidence for TCR-intrinsic specificity for MHCII. *Proc. Natl Acad. Sci. USA* **113**, 3000–3005 (2016).
101. Yin, Y., Wang, X. X. & Mariuzza, R. A. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proc. Natl Acad. Sci. USA* **109**, 5405–5410 (2012).
102. He, Y. et al. Identification of the docking site for CD3 on the T cell receptor beta chain by solution NMR. *J. Biol. Chem.* **290**, 19796–19805 (2015).
103. Li, Y., Yin, Y. & Mariuzza, R. A. Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. *Front. Immunol.* **4**, 206 (2013).
104. Barnard, D. L., Lan, M. S., Metzgar, R. S. & Finn, O. J. Specific, major histocompatibility complex-unrestricted recognition of tumor-associated mucins by human cytotoxic T cells. *Proc. Natl Acad. Sci. USA* **86**, 7159–7163 (1989).
105. Hanada, K., Wang, Q. J., Inozume, T. & Yang, J. C. Molecular identification of an MHC-independent ligand recognized by a human alpha/beta T-cell receptor. *Blood* **117**, 4816–4825 (2011).
106. Magarian-Blander, J., Ciborowski, P., Hsia, S., Watkins, S. C. & Finn, O. J. Intercellular and intracellular events following the MHC-unrestricted TCR recognition of a tumor-specific peptide epitope on the epithelial antigen MUC1. *J. Immunol.* **160**, 3111–3120 (1998).
107. Rao, A., Ko, W. W., Faas, S. J. & Cantor, H. Binding of antigen in the absence of histocompatibility proteins by arsonate-reactive T-cell clones. *Cell* **36**, 879–888 (1984).
108. Siliciano, R. F. et al. Direct evidence for the existence of nominal antigen binding sites on T cell surface Ti alpha-beta heterodimers of MHC-restricted T cell clones. *Cell* **47**, 161–171 (1986).
109. Ferreira, M. A. et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* **86**, 88–92 (2010).
110. Gulwani-Akolkar, B. et al. Do HLA genes play a prominent role in determining T cell receptor V alpha segment usage in humans? *J. Immunol.* **154**, 3843–3851 (1995).
111. Klarenbeek, P. L. et al. Somatic variation of T-cell receptor genes strongly associate with HLA class restriction. *PLoS One* **10**, e0140815 (2015).
112. Sim, B. C., Zerva, L., Greene, M. I. & Gascoigne, N. R. Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science* **273**, 963–966 (1996).
113. Rubelt, F. et al. Individual heritable differences result in unique cell lymphocyte receptor repertoires of naive and antigen-experienced cells. *Nat. Commun.* **7**, 11112 (2016).
114. Zvyagin, I. V. et al. Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. *Proc. Natl Acad. Sci. USA* **111**, 5980–5985 (2014).
115. Emerson, R. O. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* **49**, 659–665 (2017).
116. Sharon, E. et al. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* **48**, 995–1002 (2016).
117. Madi, A. et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife* **6**, e22057 (2017).
118. Dash, P. et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).
119. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017). **References 118 and 119 highlight the power of systems-based immunology in identifying predictable features in TCR responses.**
120. Mayr, E. Cause and effect in biology. *Science* **134**, 1501–1506 (1961).
121. Altman, J. D. et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
122. Ding, Y. H., Baker, B. M., Garbczci, D. N., Biddison, W. E. & Wiley, D. C. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity* **11**, 45–56 (1999).
123. Reiser, J. B. et al. Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nat. Immunol.* **1**, 291–297 (2000).
124. Degano, M. et al. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity* **12**, 251–261 (2000).
125. Moon, J. J. et al. Naive CD4⁺ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007). **This study describes tetramer-based magnetic enrichment, which enables the routine detection of antigen-specific CD4⁺ and CD8⁺ T cells from naive individuals.**
126. Robins, H. S. et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* **114**, 4099–4107 (2009).
127. Dash, P. et al. Paired analysis of TCRalpha and TCRbeta chains at the single-cell level in mice. *J. Clin. Invest.* **121**, 288–295 (2011).
128. Wang, G. C., Dash, P., McCullers, J. A., Doherty, P. C. & Thomas, P. G. T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Sci. Transl. Med.* **4**, 128ra142 (2012).
129. Kim, S. M. et al. Analysis of the paired TCR alpha- and beta-chains of single human T cells. *PLOS ONE* **7**, e37338 (2012). **References 127–129 describe the analysis of the TCR α -chain and β -chain from single cells in mice and humans.**
130. Bolotin, D. A. et al. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur. J. Immunol.* **42**, 3073–3083 (2012).
131. Bolotin, D. A. et al. MITCR: software for T-cell receptor sequencing data analysis. *Nat. Methods* **10**, 813–814 (2013).
132. Turchaninova, M. A. et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur. J. Immunol.* **43**, 2507–2515 (2013).
133. Howie, B. et al. High-throughput pairing of T cell receptor alpha and beta sequences. *Sci. Transl. Med.* **7**, 301ra131 (2015).
134. Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).

Acknowledgements

The authors thank P. Zareie for helpful comments and contributions. This work was supported by funding from the Australian National Health and Medical Research Council (NHMRC) and the Australian Research Council (ARC). N.L.L.G. is an ARC Future Fellow, S.G. is a Monash Senior Research Fellow and J.R. is an Australian ARC Laureate Fellow.

Author contributions

All authors researched data for the article, made substantial contributions to discussions of the content, wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information

Nature Reviews Immunology thanks B. Baker, C. Garcia and P. Marrack for their contribution to the peer review of this work.

Supplementary information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41577-018-0007-5>.

RELATED LINKS

Protein Data Bank (PDB): <http://www.rcsb.org/>