Understanding T-cell Receptor Signaling Through Computational Modeling and De-Novo Protein Design

William Louis White

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
David Baker, Co-chair
Hao Yuan Kueh, Co-chair
Philip H. Bradley

Program Authorized to Offer Degree:
Department of Bioengineering

University of Washington

Abstract

Understanding T-cell Receptor Signaling Through Computational Modeling and De-Novo Protein Design

William Louis White

Chairs of the Supervisory Committee:
David Baker
Department of Biochemistry
Hao Yuan Kueh
Department of Bioengineering

In order to successfully defend the host from a wide range of pathogens and cancer, T-cells must be able to recognize these threats while ignoring healthy tissues. If T-cell threat recognition is not properly calibrated, pathologies such as autoimmunity or immune deficiency can arise. The mechanisms that control T-cell recognition and activation can be broadly divided into two categories: the structural mechanism by which T-cell receptors (TCRs) can bind specifically to foreign peptides presented on the major histocompatibility complex (pMHC) but not self-pMHCs, and the biochemical mechanism by which TCR binding events are translated into T-cell activation. Here, I describe two projects that are together focused on developing tools and models to better understand each of these critical aspects of T-cell function. First, I developed a *de-novo* designed stabilizing domain which allows MHCs to be expressed solubly in *E. coli*, allowing much faster and easier production of pMHC reagents for a variety of applications in the study of pMHC-TCR interactions. Second, I developed a model of early T-cell signaling steps which explains several important aspects of TCR signaling through the clustering dynamics of signaling molecules. In combination, these two projects provide important tools and insights in the study of T-cell antigen recognition and selectivity.

Table of Contents

Introduction

T-cells serve a critical role in the immune response, detecting and destroying threats to the host organism that are contained within its own cells. They do this by scanning peptide-major histocompatibility complexes (pMHC) presented on cells throughout the body, looking for rare foreign peptides among highly abundant self-peptides[1]. In order to find these foreign pMHCs, each T-cell uses its own unique T-cell receptor (TCR) which has been selected to avoid detection of self peptides, while leaving open the possibility of detecting a specific set of foreign peptides[2]. These carefully chosen TCRs allow T-cells to detect almost any peptide that was not derived from a native host protein, ranging from viral or bacterial proteins to host proteins that have been mutated in cancer[1,3]. Their simultaneous flexibility to detect any foreign peptide, sensitivity to respond even to very low amounts of a foreign pMHC, and selectivity to ignore the large array of self peptides give T-cells the ability to efficiently eliminate threats from the body.

Given their many useful properties in sensing and eliminating infections and cancer, T-cells have attracted a lot of attention both as therapeutic targets and therapies themselves. However, in order to fully harness their potential as therapeutics, it is important to fully understand the mechanisms that underlie the qualities that make T-cells useful. Many research groups have investigated the TCR-pMHC interaction to understand which TCRs can recognize which peptides, and how specific those interactions are[4–10]. These studies reveal a high degree of complexity: while there are general trends, such as a consistently preferred docking orientation[9], often the determinants of specificity are different from one TCR or pMHC to the next[5]. Other studies have looked into the biochemical mechanisms that connect the TCR-pMHC binding event to T-cell activity; these reveal a multitude of mechanisms, including TCR clustering[11,12], mechanical forces[13,14], physical exclusion of inhibitory molecules from the vicinity of the TCR[15,16], and condensation of downstream signaling components[17–19]. These biochemical mechanisms have been incorporated into a variety of computational models that offer insight into how these mechanisms are able to achieve the remarkable properties of TCR signaling[20–29].

Despite this large body of work, our understanding of how T-cell specificity and sensitivity can stem from the structural details of the TCR-pMHC interaction and the biochemical signaling steps that occur in response to this interaction is still incomplete. The reason for this incomplete knowledge largely stems from the limitations of the tools and models that are currently available to study these phenomena. In the case of TCR-pMHC binding specificity, progress is significantly hindered by the difficult process that is necessary to produce pMHC reagents for these experiments[30]. Part of the work presented below describes the development of a *de-novo* designed protein domain that allows production of pMHC reagents through a much easier and more accessible method. In the case of computational models of TCR signaling these experimental limitations are not relevant, but most current models overlook an important feature of the signaling pathway: the formation of signaling condensates[18]. Another part of the work presented below describes the development of a model that can explain key properties of the TCR signaling as outcomes of these condensation events, and shows experimental data supporting key model predictions. Together these tools and models represent a significant advance in our ability to understand and probe specificity and sensitivity of T-cells.

Chapter 1: Designed stabilizing domain enables soluble expression of the class I MHC peptide binding groove

Abstract

The interactions between peptides displayed in the major histocompatibility complex (pMHC) and T-cell receptors play key roles in infectious disease, cancer and autoimmunity, making the mapping of these interactions central to immunobiology and therapeutics development. The large-scale mapping of pMHC-TCR interactions has been limited by misfolding and aggregation of MHCs when expressed in *E. coli*. Here, we describe the de novo design of a ~70 residue protein domain that structurally mimics the interactions made by β2-microglobulin and allows soluble expression of the peptide binding groove of class I MHCs in *E. coli*. These soluble, monomeric, antigen-receptive, truncated (SMART) MHCs retain the peptide- and TCR-binding properties of native MHCs, providing a genetically encodable alternative to streptavidin-tetramers for T-cell staining, and enabling high throughput mapping of pMHC-TCR interactions.

Introduction

Recombinantly expressed soluble peptide-major histocompatibility complexes (pMHCs) are widely used as a staining reagent to identify or isolate the T-cell subset that can respond to a peptide of interest[31], and have led to important discoveries in T-cell specificity[6,8,32], infectious disease[33], autoimmunity[34,35], cancer immunology[36–38] and many other areas. Scaling up of these studies has been limited by the pMHC production process, which involves separate expression in *E. coli* of each of the two MHC chains as insoluble inclusion bodies, solubilization, and subsequent refolding in the presence of the desired peptide[39], which is expensive, slow, and inefficient. To reduce the burden of refolding, methods have been developed by which a single refolding reaction can be split and loaded with many different peptides[30,40–42]. However these methods are still limited by the difficulty of refolding. Other methods have used eukaryotic expression systems, such as viral[7,43] or yeast display[4,44] libraries, that can produce a well-folded MHC structure natively[9], but these approaches can be limited by library size or expression levels.

A more stable MHC molecule that could be readily expressed in *E. coli* or yeast would enable the study of peptide-specific T-cell populations at an unprecedented scale. Instead of relying on specialized facilities to produce refolded pMHCs for staining experiments[39], immunologists could produce them in-house, dramatically improving their ability to iterate through multiple peptide variants or MHC alleles. Similarly, a stable MHC molecule could facilitate screening of large peptide libraries by yeast display without the need for prior optimization of the MHC sequence for display. We reasoned that such a molecule could be created by leveraging the stability of *de-novo* designed proteins and recent advances in protein-protein interface design[45] to replace a portion of the MHC molecule with a designed protein scaffold. This scaffold would replace or improve on the native interactions that maintain the structure of the MHC protein while preserving the peptide and TCR binding properties of the stabilized MHC. We set out to use this approach to design a soluble, monomeric, antigen-receptive, truncated (SMART) MHC system.

Results

Design of SMART MHCs

The structure of a native class I MHC is composed of four domains split between two chains: the α1, α2, and α3 domains constitute the MHC chain which heterodimerizes with β2 microglobulin (β2m) to form the full MHC structure[46]. The α1 and α2 domains consist of a pair of α-helices lying over a β-sheet that forms the peptide binding groove[46] and are stabilized by interactions with β2m on the opposite side of the β-sheet. Following the precedent of previously described truncated "mini-MHC" systems, we removed the α3 and β2m domains completely[5,47,48] (fig. 1B, left), which leaves an exposed hydrophobic patch on the underside of the β-sheet. We used computational methods previously developed for the design of protein binders for arbitrary target proteins[45] to design small protein domains to bind this hydrophobic patch and stabilize the structure. First, we docked small *de-novo* designed helical bundles (fig. 1B, left) against the bottom of the β-sheet of a truncated MHC allele (mouse H-2D^b, which we chose as a representative example as all MHCs are very similar structurally). Docking positions were selected such that the backbone positions of the bundle could accommodate side chains that made favorable contacts with the MHC, in some cases directly replacing native side chains of β2m. We then designed the bundle sequences to make these contacts and stabilize the MHC fold (fig. 1B, middle). Finally we linked the designed stabilizing domains to the N-terminus of the truncated MHC by a flexible linker (fig. 1A, right).

We screened roughly 10^4 designs using yeast surface display[49], sorting first for display on the yeast surface (fig. S1A), and then for binding to fluorophore-labeled gp33 peptide (fig. S1B), which is strongly bound by native H-2D^b [50]. Deep sequencing of these populations revealed that designs which contained a tryptophan residue placed similarly to W60 of β2m in the native structure were much more likely to be positively enriched, consistent with the finding that mutations at W60 can significantly destabilize the interaction between β2m and the MHC[51,52]. We tested 29 enriched designs for soluble expression in *E. coli*; the most effective stabilizing domain (hit6) preserved MHC folding in both yeast and *E. coli* expression systems, but with low yield (fig. 1F, left). We identified likely cleavage sites based on the masses of the proteolytic fragments, and then redesigned the sequence near those sites (fig. 1E, top), preventing mutations in amino acids that directly interact with the peptide of TCR, and favoring substitutions commonly observed in other MHC alleles. We selected the variant (CSM8 MHC) which best prevented cleavage, and used machine learning methods[53–55] to design a more structured linker to replace the original GS linker (fig. 1E, bottom), and selected the design that produced the largest improvement in soluble monomeric yield (fig. 1F, right). This design could be solubly expressed not only in E coli but also in cell free protein synthesis (CFPS)[56] with much higher yields than the native MHC (fig. 1G). We refer to this design with the CSM8 mutations and optimized linker as SMART MHC below.

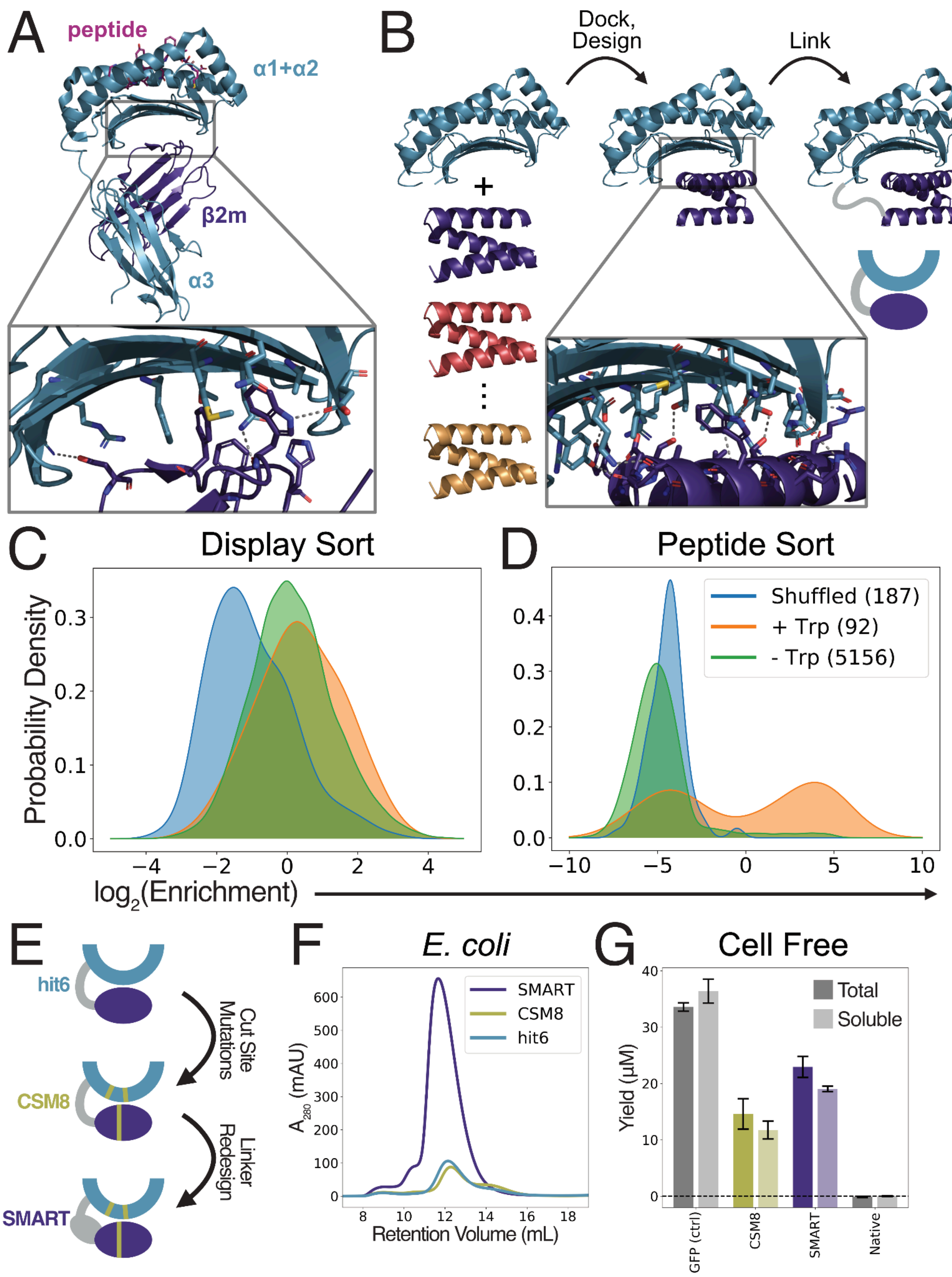

Figure 1. Design and screening of SMART MHCs. A) Structure of native H-2D^b in complex with the gp33 peptide (PDB ID: 1S7U). Inset shows the key interactions that β2m makes with

the α1 and α2 domains. B) Schematic of the initial design process used to generate a library of candidate stabilizing domains showing the truncated H-2D^b with candidate stabilizers (left), the placement of the stabilizing domain (middle), and the final linked version used for yeast display (right). C,D) KDEs of the distribution of enrichment values of all designs is either the expression (C) or peptide binding (D) yeast display sorts. Designs are split into three categories for KDE calculations: sequence-shuffled negative controls (shuffled), designs that do not replace W60 from β2m (-Trp), and designs that do replace W60 from β2m (+Trp). Relative enrichment for each design was calculated by dividing its abundance after the indicated sort by its abundance prior to that sort. E) Schematic of the iterative improvements made to hit6 showing the cleavage site mutations and redesigned linker. F) SEC traces from an S75 column for soluble material from 50mL *E. coli* expression cultures of each iterative improvement to hit6. G) Bar plot showing yields from cell free protein synthesis reactions for GFP (positive control) and several H-2D^b variants.

SMART H-2D^b retains native binding properties and structure

To verify that the soluble material we produced in *E. coli* retained the functional characteristics of the native MHC, we first measured the binding affinity of SMART H-2D^b for the gp33 peptide using fluorescence polarization (FP) measurements. Because SMART MHCs can be produced without a place-holder peptide, we directly measured binding by titrating empty SMART MHC into a constant concentration of peptide, obtaining an affinity of 884±14pM (fig. 2A), stronger than the reported value of 5nM for native H-2D^b [57]. We also found that SMART H-2D^b was highly stable, retaining this tight binding affinity after storage for at least one month at 4^{o}C, and after one freeze-thaw cycle (fig. 2A).

Given that SMART H-2D^b can bind the gp33 peptide, we next tested whether it was able to do so in a manner that preserved the interactions with a relevant TCR. We used the P14 TCR, which binds H-2D^b in complex with the gp33 peptide[50,58]. We first loaded the gp33 peptide (or a variant thereof) into CSM8 H-2D^b and then measured the binding affinity of this complex to the P14 TCR using surface plasmon resonance (SPR). The affinities were very similar to those of native H-2D^b peptide complexes (table 1; fig. S2; fig. 2C), indicating that the designed domain promotes folding to and stabilizes the TCR-binding conformation of the MHC.

Table 1. P14 TCR Binding Affinities of gp33 Peptide Variants. All native affinities were measured previously by Duru *et al.*[58]

Peptide name	Peptide sequence	CSM8 H-2D^b affinity	Native H-2D^b affinity
gp33	KAVYNFATM	14.1μM	8.6μM
V3P-Y4F	KAPFNFATM	75.6μM	80μM
Y4F	KAVFNFATM	609μM	No binding detected

We additionally determined the crystal structure of CSM8 H-2D^{b} in complex with the gp33 peptide. This structure shows excellent alignment with both the native structure[59] (fig. 2D) and the design model (fig. 2E), with CA-RMSDs of 0.39 Å and 0.53 Å, respectively. The peptide backbone is very well aligned with the native conformation, and the side chains show only relatively minor changes between the two structures (fig. 2G). These small changes, along with a few slight shifts elsewhere in the structure could explain the slight deviations in peptide binding affinities we observed in our SPR experiments. The crystal structure also shows a tryptophan side chain replacing the critical interactions provided by W60 of β2m, as designed (fig. 2F). Overall, these results demonstrate that SMART H-2D^{b} is able to retain all of the necessary structural features that allow peptide and TCR binding, without the refolding requirements of the native MHC.

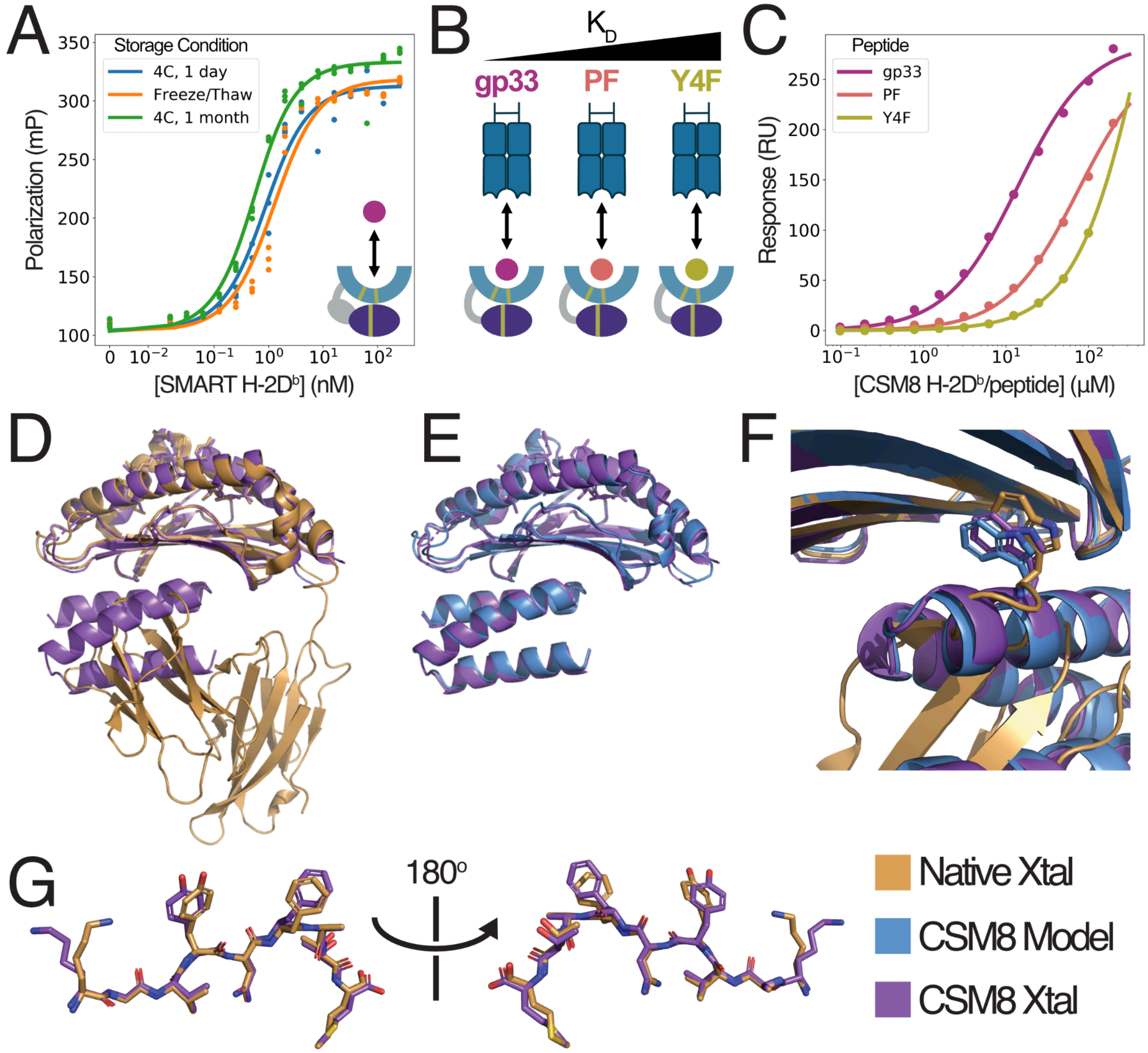

Figure 2. SMART H-2D^{b} retains native interactions and structure. A) FP data (filled circles) and fitted binding curves (lines) for FITC-gp33 binding to purified SMART H-2D^{b} that was either

stored at 4°C for one day (4C, 1 day) or 1 month (4C, 1 month), or flash-frozen and stored at -80°C for one day and thawed immediately before use (Freeze/Thaw). B) Cartoon representations of gp33 peptide variants with varying affinity for the P14 TCR. C) Equilibrium binding measurements (filled circles) and fitted binding curves (lines) for complexes of CSM8 H-2D^b with three different peptides binding to immobilized P14 TCR. D) Structural alignment of the CSM8 H-2D^b/gp33 complex crystal structure (purple) to that of the native (gold). E) Structural alignment of the CSM8 H-2D^b/gp33 complex crystal structure (purple) to the design model (blue). F) Structural alignment of all structures from D and E highlighting the positioning of W60 in the native structure relative to W14 of the stabilizing domain. G) Comparison of the peptide conformation for the structures in D.

*Designed stabilizing domain allows soluble expression of functional HLA A*02:01*

Next, we tested the ability of our stabilizing domain to generalize to other MHC alleles. Based on the structural and sequence similarity of H-2D^b to a variety of human MHCs, we did not make any modifications to the stabilizing domain, instead varying only the MHC sequence. We selected HLA A*02:01 to test initially due to its abundance in the human population, and because there are many well-characterized peptides and TCRs that bind to it[60,61]. We found that CSM8 A*02:01 could be expressed solubly in *E. coli*, with roughly half of the material produced in a monomic and half in a dimeric (and likely misfolded) state (fig. 3A). For all of the following characterization we used only the monomeric fraction, as purified by SEC.

To test the ability of CSM8 A*02:01 to present peptides, we selected the NY-ESO-1 peptide because of its relevance in cancer[62] and the existence of a well-characterized TCR that binds the NY-ESO-1/A*02:01 complex[61]. Using FP methods very similar to those used for SMART H-2D^b, we measured the binding affinity of CSM8 A*02:01 to NY-ESO-1 to be roughly 190nM (fig. 3B), similar to the 40nM affinity measured for the native A*02:01[63]. The difference in affinity between the native and CSM8 A*02:01 could be due to partial unfolding of the CSM8 A*02:01, or to contamination of the monomeric species with dimer due to imperfect separation by SEC. Overall this difference is relatively minor, and the FP measurements indicate that peptide binding is maintained with CSM8 A*02:01.

Given that CSM8 A*02:01 can bind the NY-ESO-1 peptide, we next checked whether it also retained binding to the 1G4 TCR[62]. Using similar SPR methods to those used for H-2D^b we determined the affinity of the NY-ESO-1/CSM8 A*02:01 complex to the 1G4 TCR with two variants of the peptide (fig. 3C). We found that the measured affinities were roughly 7-9 fold weaker than the same values for native A*02:01 (table 2). Despite this difference in absolute binding affinity the ranking of the affinities for peptide variants was maintained, indicating that, although some of the CSM8 A*02:01 might be improperly folded, the proportion that is folded can be recognized by a relevant TCR. Overall, these data suggest that, although it is less effective, our stabilizing domain can stabilize an allele other than the one it was designed for.

Table 2. 1G4 TCR Binding Affinities of NY-ESO-1 Peptide Variants. All native affinities were measured previously by Chen *et al*[61]*.*

Peptide name	Peptide sequence	CSM8 A*02:01 affinity	Native A*02:01 affinity
9V	SLLMWITQV	38.9μM	5.7μM
9C	SLLMWITQC	123μM	13.3μM

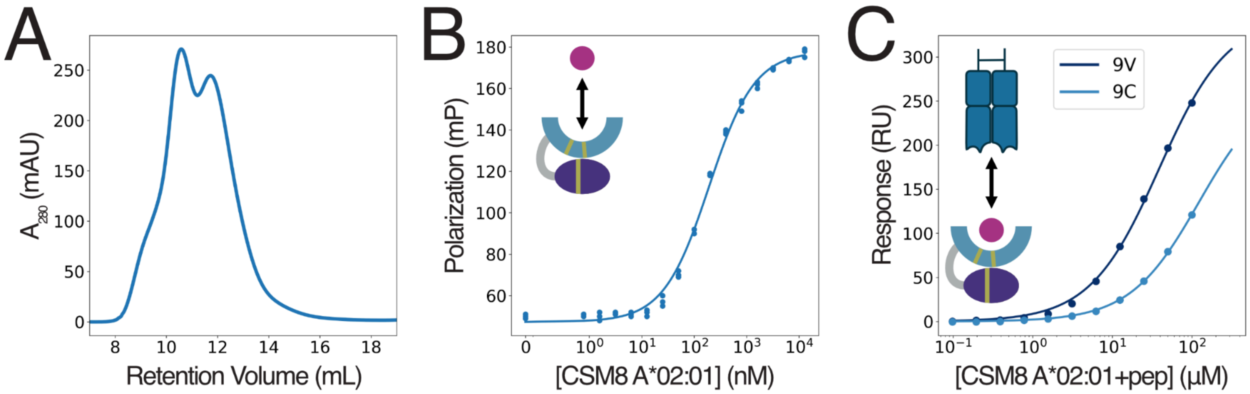

Figure 3. CSM8 A*02:01 retains native interactions. A) SEC trace for empty CSM8 A*02:01. Elution volume for the monomeric species is roughly 12mL. B) FP data (filled circles) and fitted binding curve (line) for CSM8 A*02:01 binding to AF488-NY-ESO-1 peptide. C) Equilibrium binding data (filled circles) and fitted binding curves (lines) from SPR experiments for complexes of CSM8 A*02:01 with variants of the NY-ESO-1 peptide binding to immobilized 1G4 TCR.

*Stabilizing domain improves yeast display of HLA A*02:01*

As previously mentioned, yeast display can be a powerful tool to screen libraries of peptide variants for TCR binding[4,44]. We again used A*02:01 as a test case, this time using the TAX peptide and A6 TCR that recognizes the A*02:01/TAX complex[60]. The single-chain SMART A*02:01/TAX complex was displayed on the C-terminus of Aga2 using a mutant A*02:01 (W167A) to allow the peptide linker to leave the peptide binding groove. We found that SMART A*02:01 is displayed at much higher levels than native A*02:01, regardless of whether a peptide is fused to it (fig. 4A). The elevated expression levels also lead to an increase in binding of an A6 TCR tetramer when the TAX peptide was present (fig. 4B). These results indicate that the SMART system can significantly improve yeast display of MHCs, opening the door to generation of very large libraries of peptide-MHC allele combinations for large scale mapping of pMHC-TCR interactions.

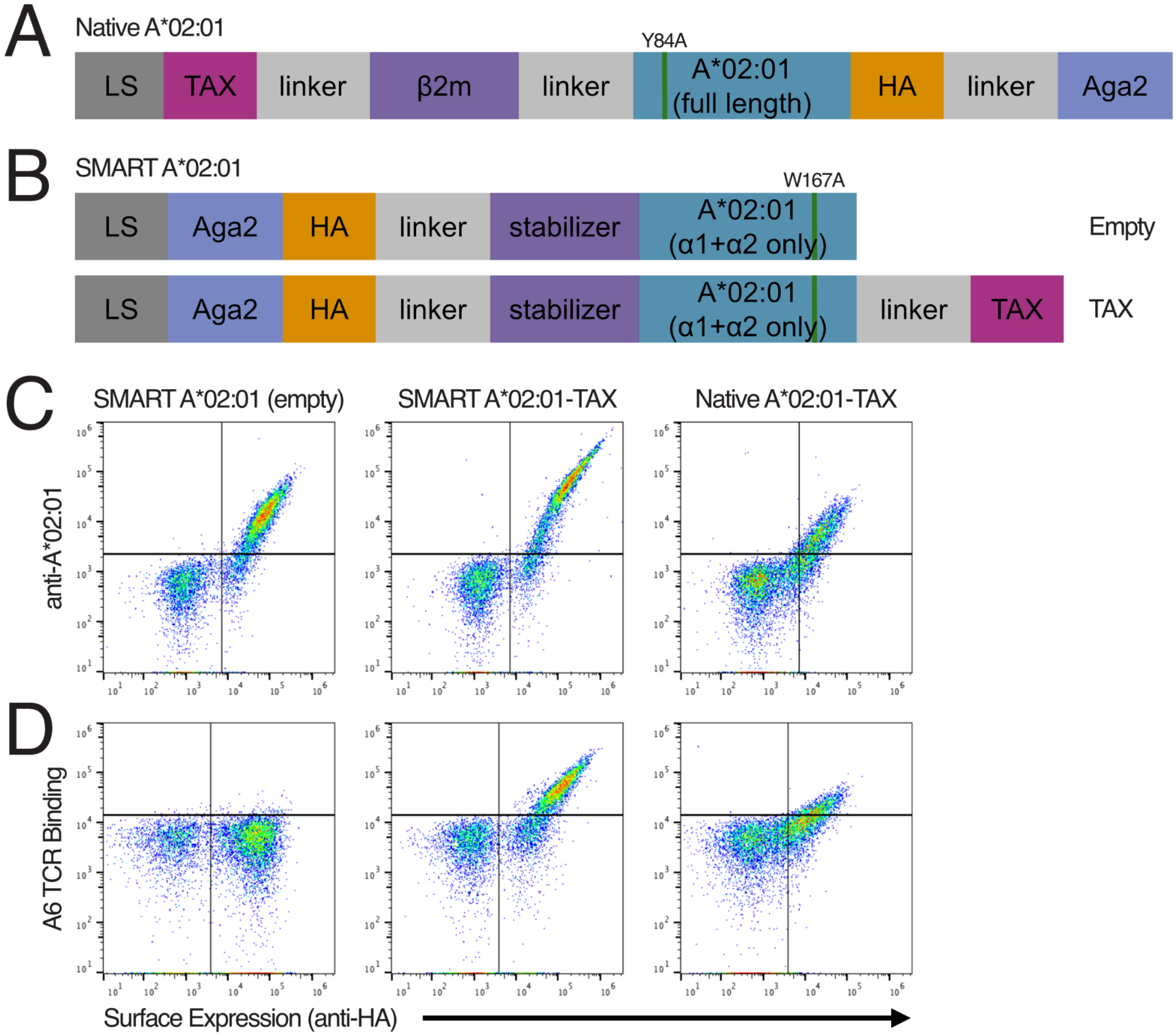

Figure 4. SMART A*02:01 improves pMHC yeast display. A) Schematic of native A*02:01 yeast display construct. The Y84A mutation is used to allow room for the linker on the C-terminus of the peptide to leave the binding pocket. B) Schematic of SMART A*02:01 display constructs with and without TAX peptide fused. The W167A mutation is used to allow room for the linker on the N-terminus of the peptide to leave the binding pocket. LS: leader sequence for surface display. HA: HA peptide tag for measuring surface display. C) Scatterplots from flow cytometry measurements of expression levels of various A*02:01 constructs on the surface of yeast. Surface expression was measured by binding of an anti-HA antibody, and by binding of an A*02:01-specific antibody. D) Scatterplots from flow cytometry measurements of A6 TCR binding of various A*02:01 constructs on the surface of yeast. TCR binding was measured with streptavidin tetramers of the A6 TCR.

Peptide-fused SMART MHC oligomers stain T-cells in a TCR-specific manner

A major application of pMHCs in immunological research is the staining of T-cells with pMHC tetramers[31]. We therefore tested whether SMART MHCs could be converted into a similar multimeric staining reagent. Rather than using streptavidin to tetramerize the SMART MHCs, as is typically done, we chose to directly fuse them to a *de-novo* designed oligomeric protein which assembles into a tetrahedral symmetry containing 12 subunits, bypassing the biotinylation step which is necessary for streptavidin-based tetramerization. We fused the peptide to be displayed to the N-terminus of the SMART MHC along with a SUMO tag, incorporating the Y84A MHC mutation to allow the peptide linker to exit the peptide binding groove. Co-expression of the Ulp1 protease allows the N-terminus of the peptide to be cleanly cleaved, allowing it to bind properly in the peptide binding groove. On the C-terminus of the oligomer, we fused a Myc tag to allow for antibody staining (fig. 5A).

To assess the ability of the SMART MHC oligomers to stain T-cells, we mixed two populations of T-cells: one expressing the P14 TCR and the other expressing an unrelated TCR along with mTagBFP to allow the two cell types to be distinguished independently of TCR staining. We then made SMART H-2D^b oligomers fused to several variants of the gp33 peptide with known affinities to the P14 TCR[50,58] and assessed their ability to stain the T-cell mixture. We found that peptides with affinities similar to unmutated gp33 (V3P and M9C) showed bright staining of the P14 T-cells relative to the control cells, with the highest affinity variant, V3P, achieving the brightest staining. In contrast, we were unable to detect binding to gp33 variants with significantly reduced affinities (V3P+Y4F and Y4F; staining is somewhat less sensitive than the SPR methods used to measure the binding affinities). Overall, the success of the SMART MHC oligomers in staining T-cells in both a TCR- and peptide- specific manner indicates that they can be used in place of pMHC tetramers without the need for refolding and biotinylation.

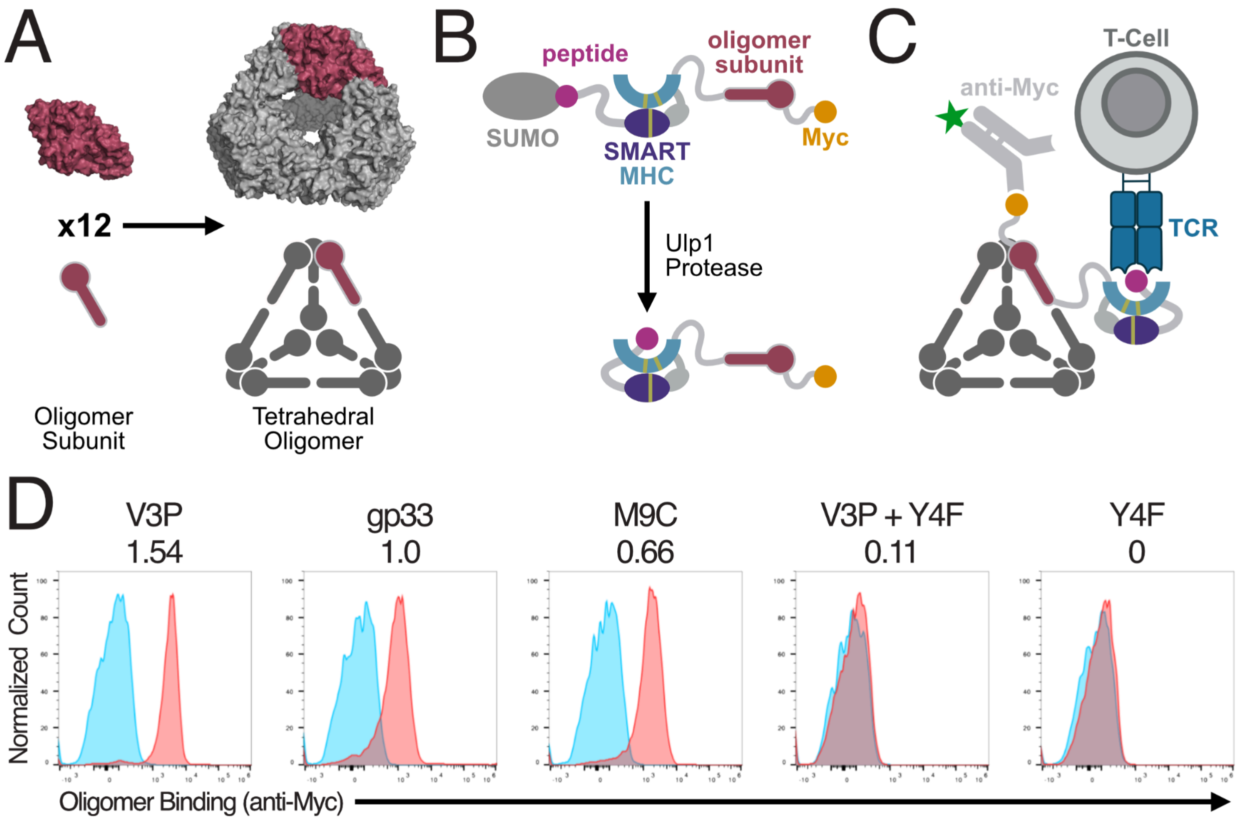

Figure 5. Peptide-fused SMART MHC oligomers stain T-cells in a TCR-specific manner. A) Twelve identical subunits assemble to form a tetrahedral oligomer. A single subunit is highlighted in red. B) Schematic of the SMART MHC oligomer sequence. Ulp1 protease is used to cleave the SUMO tag from the peptide, leaving a clean peptide N-terminus to bind the SMART MHC peptide binding groove. The MHC sequence bears the Y84A mutation to allow the C-terminal linker from the peptide to leave the binding groove. The peptide-fused SMART MHC is then fused to an oligomerization domain, followed by a Myc tag. C) The SMART MHC oligomer self-assembles into the same shape as the base oligomer in (A). Each subunit presents a peptide-fused SMART MHC to the T-cell and a Myc tag for secondary staining, but only one subunit is shown with these for simplicity. D) Flow cytometry measurements of T-cell staining of P14 T-cells (red) or control TCR T-cells (blue) with SMART H-2D^b oligomers fused to several variants of the gp33 peptide. Numerical values under the peptide names indicate the binding strength of each peptide in the native MHC context for the P14 TCR relative to gp33 ($K_{D,gp33}/K_{D,mut}$). Binding strength values for V3P, V3P+Y4F, and Y4F were previously published by Duru *et al.*[58]. Value for M9C was published previously by Boulter *et al.*[50]

Discussion

Our success in creating a single stabilizing domain that allows soluble expression in *E. coli* of both the H-2D^b and A*02:01 MHC alleles paves the way to soluble and high throughput expression for all MHC alleles and peptide-MHC pairs. While the current SMART system does not rescue soluble expression of all MHC-peptide pairs, we are optimistic that with direct experimental optimization for expression and display it should be broadly applicable. While we focused on class I MHC in this work, our approach should be readily extensible to class II alleles. The ability to produce fully genetically encoded tetramer-like T-cell staining reagent without the need for inefficient and labor-intensive refolding protocols should very substantially increase the number of peptide-MHC combinations that can be tested in T-cell tracking and sorting experiments. Such efforts could have wide impact across immunology, ranging from improved tracking of immune responses to infectious disease and identification of cancer-targeting T-cell clones. Similarly, efficient display of SMART pMHCs on yeast will be very useful for probing the determinants of pMHC-TCR interactions. For example, a single yeast library displaying SMART versions of many MHC alleles, each in combination with a wide range of peptide sequences, could be produced and sorted to rapidly and comprehensively map the binding specificity of any TCR of interest. More generally, SMART MHCs have the potential to rapidly accelerate our understanding of T-cell behavior and TCR specificity.

Methods

Stabilizer library design

We used previously developed computational methods to design protein binders for arbitrary target proteins[45] to design this stabilizing domain. The “target” supplied to this method was the α1 and α2 domains of the H-2D^b structure (PDB: 1S7U). We ran two versions of the design protocol. The first version was a completely *de-novo* approach which allowed any amino acid to be placed anywhere near the bottom of the H-2D^b peptide binding groove in order to make a favorable interaction. The second approach specifically focused on making designs which placed amino acids in locations that allow them to replace the exact interactions that β2m makes in the native structure. Designs from both of these protocols were pooled and filtered for the quality of the stabilizer and the interactions it made with the MHC, as previously described[45]. We also included a set of negative control designs which were made by randomly scrambling the sequence (while conserving the pattern of hydrophobic and hydrophilic residues) of a random subset of the designs.

Yeast display screening of stabilizing domains

Yeast display was performed as previously described[45] with the following changes. Rather than screening our designs for binding to the α1 and α2 domains of H-2D^b, we fused the designs to those domains using a flexible poly-GS linker and screened them for surface display and binding to a FITC-labeled gp33 peptide (KAVYNFATM), with FITC linked to the amine on the lysine sidechain. Sorted populations were subsequently cultured and plasmid DNA was extracted for sequencing.

Protein expression and purification

Genes encoding the designed protein sequences were synthesized and cloned into modified pET-29b(+) E. coli plasmid expression vectors with a 6xHis tag added to the N-terminus (for monomeric versions) or the C-terminus (for oligomeric/peptide-fused versions). Plasmids were transformed into chemically competent E. coli BL21 (DE3) cells (NEB). E. coli cells were grown either in TBII-AIM overnight at 37°C (autoinduction media[64] made with TBII) or in LB medium at 37 °C until the cell density reached 1.0 at OD600, followed by addition of 1mM IPTG and overnight growth at 16°C. The cells were collected by spinning at 4,000g for 5min and then resuspended in lysis buffer (150 mM NaCl, 25mM Tris-HCL (pH 8.0), 25mM imidazole, 1mM PMSF, and 5% glycerol) with RNase. The cells were lysed with a Qsonica Sonicators sonicator for 7 min in total (3.5 min each time, 10 s on, 10 s off) with an amplitude of 80%. The soluble fraction was clarified by centrifugation at 14,000g for 30 min. The soluble fraction was purified by immobilized metal affinity chromatography (Qiagen) followed by FPLC SEC on a Superdex 75 10/300 GL column (GE Healthcare) for monomeric versions, and a Superose 6 10/300 GL column (GE Healthcare) for oligomeric versions. All protein samples were characterized by SDS–PAGE, and purity was greater than 95%. Protein concentrations were determined by absorbance at 280 nm measured with a NanoDrop spectrophotometer (Thermo Scientific) using predicted extinction coefficients.

Cleavage site mutation design

Likely cleavage sites were identified by LCMS analysis of the cleavage fragments of hit6 *E. coli* expressions. Residues near these positions that were not likely to interact with a peptide or TCR were identified by examining the structure, and Rosetta design was used to make mutations at those positions. The design protocol was restricted based on a local position-specific substitution matrix (PSSM) created from a multiple-sequence alignment of MHC sequences.

The PSSM for each residue was based only on a 9-amino acid window centered on the target position, and was restricted to MHC sequences that shared similar 3D structures in that local region, thus restricting the mutations to those that are well-represented across MHC alleles.

Linker design

Linker design was used to replace the flexible GS linker that was used in the yeast display screening with a more structured and shorter linker. Using the design model of the CSM8 variant as a starting point, we used previously developed “inpainting” methods[53] to fill in a small segment of protein structure to bridge the gap between the C-terminus of the stabilizer and N-terminus of the MHC. Sequences for the protein backbones produced by this method were designed using ProteinMPNN[54], restricted to changing only the amino acids in the “inpainted” structure. Finally, the resulting designs were evaluated using AlphaFold[55] predictions where the MHC structure was provided as a template, but the stabilizer and linker were not. Designs with high overall pLDDT scores and low PAE scores for residues in the MHC/stabilizer interface were selected for experimental testing.

Peptide binding affinity measurements

Three technical replicates of varying concentrations of SMART MHC were mixed with a constant concentration of fluorophore-labeled peptide (300pM FITC-gp33 for H-2D^b and 10nM AF488-NY-ESO-1 for A*02:01) and incubated overnight at room temperature to allow equilibration. Fluorescence polarization measurements of these samples were made with a Synergy Neo2 plate reader (BioTek instruments) with a 485/530 FP filter. Binding curves (using the non-simplified equilibrium binding equation[65]) were fitted separately to each of the triplicate measurements and averaged to determine the K_D.

TCR binding affinity measurements

TCR binding affinities were measured as previously described[58], using SMART MHCs as the mobile phase instead of native MHCs. All measurements were performed on a BIAcore T200 (GE Healthcare) at 4°C in the buffer containing 10 mM HEPES pH7.4, 150 mM NaCl, 0.005% Tween-20, 3 mM EDTA. Soluble P14-his6 was noncovalently coupled to the anti-his antibody, immobilized on a CM5-chip via standard amine coupling, and around 4000 response units of anti-his antibody was coupled, immobilizing around 1000 response units of P14-his5 (0.75 uM). A control surface was generated the same way, and up to 100 μM or 200 μM of freshly produced WWDb/peptide complexes (2-fold dilutions from stock, at least 10 concentrations in duplicate) were injected over the chip surfaces at 30 μL/mL. The sample rack was cooled to 4°C during the run. Chip surfaces were regenerated using 0.1 M Glycine-HCl pH 2.5, 500 mM NaCl, Tween 0.05% at 30 μL/min after each injection. The WWDb/peptide was injected over a control surface, and the final signal was calculated by subtracting the signal obtained on the control surface from the signal on the TCR-coupling surface, to remove the contributions of the bulk effect and possible non-specific binding. The data were then analyzed using BIAevaluation 3.0 software. The KD values were obtained from steady-state fitting of equilibrium-binding curves.

Crystallography

Crystallization was performed using the sitting-drop vapor diffusion method at 293.15 K. Drops were set up using 0.15 µl of the SMART H-2D^b/gp33 and 0.15 µl of the reservoir solution (0.2 M sodium acetate trihydrate, 0.1 M TRIS hydrochloride pH 8.5, and 30% w/v polyethylene glycol, PEG 4,000), equilibrated against 50 µl of reservoir solution. Crystals appeared after 6~13 days and were cryoprotected with a solution containing an additional 7.5% w/v PEG 4,000 in the reservoir solution and harvested using mounted CryoLoops (Hampton Research). Subsequently, the crystals were flash-frozen in liquid nitrogen and transported to the beamline for data collection. Diffraction images were collected in the automatic beamline ID30 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Diffraction data were processed using autoPROC[66]. The crystal structure was determined by molecular replacement utilizing Phaser-MR in PHENIX[67] with the design model employed as the search model. Subsequent refinement was carried out using PHENIX. Manual model building was conducted using Coot[68], and the resulting model was further refined using PHENIX[69–71]. The figures were generated using PyMOL Molecular Graphics System (Schrödinger). The final coordinates/structure factors will be deposited in the protein data bank.

Peptide-fused yeast display

Yeast display of peptide-fused SMART A*02:01 was performed as previously described[4]. In brief, 50ng pCT or pYAL plasmids encoding corresponding full length or SMART A02 constructs with TAX were electroporated into competent EBY100. The EBY100 was cultured in YPD medium for 1 hour at 30°C, spun down and continued to grow in SDCAA medium for 48 hours before induction in SGCAA for 48 hours. Display levels were evaluated with fluorophore-conjugated anti-HA and anti-A*02:01 (clone BB7.2) antibodies, and TCR binding was evaluated with TCR tetramers made by combining soluble biotinylated A6 TCR with fluorophore-conjugated streptavidin.

T-cell staining

Two Jurkat T-cell lines were used: one expressing the P14 TCR, and the other expressing an unrelated TCR (a3a TCR, recognizing the MAGE-A3/A*01:01 complex) as well as mTagBFP. Both cell lines were mixed in equal numbers and resuspended to 1M cells/mL in 100uL of staining solution (500nM SMART MHC oligomer, 25mM Tris-HCl pH 8.0, 150mM NaCl, and 5% glycerol) and incubated at 4°C for 30min. Cells were then washed twice with 100uL of Fc block (HBH (HBSS with 0.5% BSA, and 10mM HEPES) supplemented with 10% 2.4G2 cell culture supernatant). Cells were then resuspended in 50uL AF647-anti-Myc antibody diluted in Fc block and incubated at 4C for 20min. Cells were then washed twice with 100uL of HBH and resuspended to a final volume of 150uL of HBH for analysis on an Attune Nxt flow cytometer. Cells were gated into P14 (mTagBFP-) or a3a (mTagBFP+) populations and the brightness of the AF647 stain for each population was compared.

Supplemental Figures

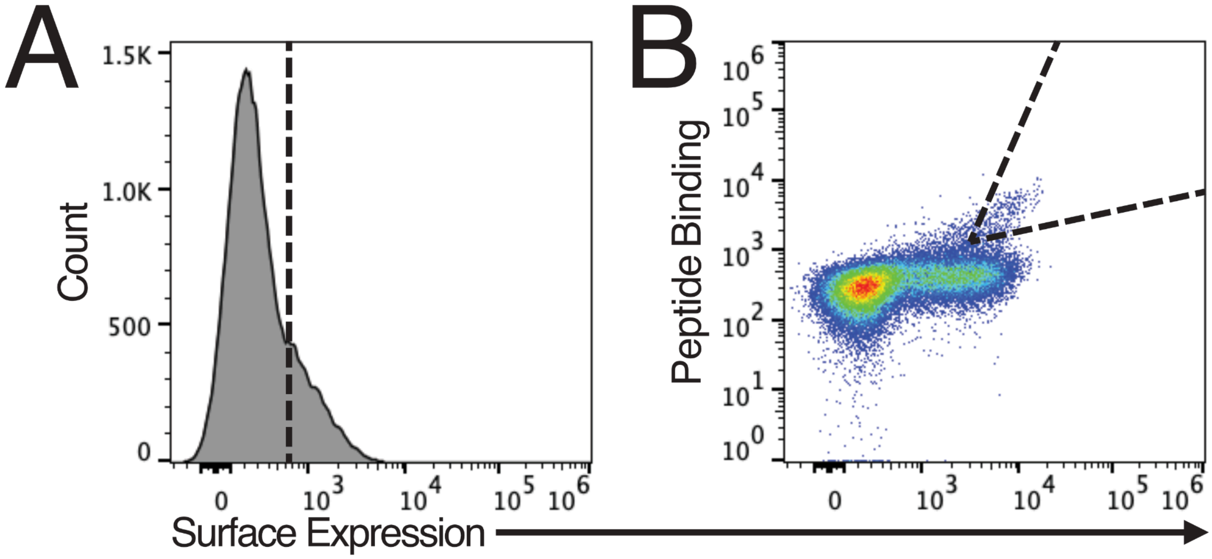

Figure S1. Yeast display sorting of the stabilizer library. A) Histogram of surface expression levels of the stabilizer library. Surface expression was evaluated by antibody staining of a C-terminal Myc tag. Cells with high surface expression were collected after the sort and recultured for the next sort. B) Joint distribution of the surface expression and peptide binding levels of the stabilizer library after the sort in A. Peptide binding was evaluated with FITC-labeled gp33 peptide. Cells with high surface expression and peptide binding were collected after the sort.

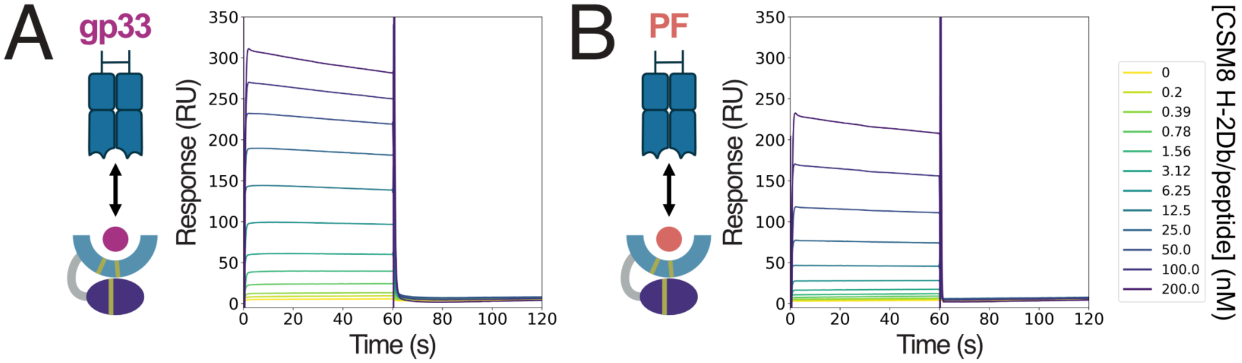

Figure S2. P14 TCR binding kinetics. Representative SPR traces from TCR pounding experiments with the P14 TCR and several variants of the gp33 peptide complexed to CSM8 H-2D^b. A) Data from the gp33 peptide. B) Data from the PF variant of the gp33 peptide.

Chapter 2: Proofreading and single-molecule sensitivity in T-cell receptor signaling by condensate nucleation

Abstract

T-cells display the remarkable ability to detect single foreign peptides displayed on target cells, while ignoring highly abundant self peptides. This selectivity has been explained by kinetic proofreading in the T-cell receptor (TCR) signaling pathway, which prevents responses to short-lived binding events regardless of their abundance. However, the biochemical mechanisms that drive kinetic proofreading have remained unclear. Here, using computational modeling, we show that these key signaling properties of the TCR pathway can emerge from the dynamics of LAT phosphorylation, diffusion, and condensation following TCR-pMHC binding. In this model, time delays in LAT condensate nucleation underlie kinetic proofreading, enabling selective signaling responses to high-affinity pMHC ligands. The cooperativity in the nucleation and growth of LAT condensates also provides a mechanism to amplify weak signals from single foreign peptides and for condensates to grow with increasing antigen numbers. In contrast to other models, condensate-nucleation proofreading predicts a dependence of signal strength on pMHC spacing at fixed number, a prediction we validated experimentally using a protein scaffold to present pMHCs at defined intervals. Our results suggest that nucleation-condensation proofreading underlies the remarkable antigen detection capabilities of the TCR signaling pathway.

Introduction

The TCR signaling pathway simultaneously achieves several functional properties that enable T-cells to sense and appropriately respond to challenges of varying types and severities[72]. Firstly, it senses foreign peptide antigens displayed on major histocompatibility complex ligands (pMHCs) in a highly sensitive manner, in some cases detecting single copies of foreign pMHCs[1]. Secondly, it is highly selective so as to avoid responding to more abundant self antigens, which have a weaker binding affinity for the TCR[32]. And finally, although the TCR signaling pathway activates in a digital, all-or-none manner upon pMHC recognition[20,73,74], it can also tune its response to different levels of antigen input in an analog manner, enabling T-cells to tailor their responses to the magnitude of the threat[21,75]. Despite significant advances in our knowledge of the biochemical steps involved in TCR signaling[76,77], it remains incompletely understood how the TCR signaling pathway is able to simultaneously achieve sensitivity, selectivity, and dynamic range in antigen sensing.

The first step in T-cell activation after a pMHC binds to the TCR is phosphorylation of the ITAM-containing cytoplasmic tails of the TCR-associated CD3 proteins by Lck[76]. The kinase ZAP70 can then bind to the phospho-ITAMs and become activated through phosphorylation by Lck and by trans-autophosphorylation[78], and can in turn phosphorylate a variety of downstream substrates including the transmembrane protein LAT[76], leading to signaling pathway activation. To explain how the T-cell signaling pathway is able to selectively respond to foreign peptides based on their stronger binding affinity, the concept of kinetic proofreading has been proposed[22] and extensively studied using mathematical modeling[20,21,23–26]. In its most basic form, kinetic proofreading consists of a series of biochemical steps that can occur only when a TCR is bound to a pMHC, and that are quickly reversed when the pMHC unbinds[22]. This allows the T-cell to respond strongly to foreign antigens that bind long enough to complete all necessary biochemical steps, while ignoring self antigens that bind only briefly.

A number of extensions of the basic kinetic proofreading model have been evaluated, and shown to improve different facets of the signaling behavior. The addition of a negative feedback loop can dramatically enhance signaling selectivity[23], whereas positive feedback loops can enhance sensitivity to low pMHC copy numbers[20,27], generating an all-or-none ‘digital’ response. On the other hand, incoherent feedforward loops can expand the dynamic range of the T-cell response[21]. Together these models demonstrate each of the three critical signaling properties of T-cells, but individually each model is only able to achieve one or two. Additionally, most current models do not account for the inherent stochasticity in signaling which can result in responses that differ significantly from deterministic predictions[26,28], and can dramatically reduce selectivity[29].

The condensation and clustering of signaling molecules downstream of the TCR could constitute important events for recapitulating the cardinal properties of T cell signaling. Pioneering studies from Su and co-workers established that LAT can undergo condensation when phosphorylated. Phospho-tyrosine (pY) sites on phospho-LAT (pLAT) act as binding sites for SH2 domains in Grb2 and PLCγ, among other proteins[18,17]. Both Grb2 and PLCγ also contain SH3 domains that can bind to the proline-rich repeat regions of SOS[18,17,19]. This network of polyvalent interactions causes these molecules to form condensates tethered to the membrane by LAT[18,17,19], and recruits a variety of other molecules which act together with PLCγ and SOS to transmit the activation signal into the T-cell[76,79]. These clustering events have been overlooked by existing kinetic proofreading models, which generally assume that the reaction system is well-mixed, even though it has been determined experimentally that this is not the case[18,17,19,80]. More recent work has explored the condensation of LAT with Grb2, SOS, and PLCγ using coarse grained molecular dynamics simulations[17,81]. However, these studies focus only on the formation of the condensate and not its context within the larger signaling pathway.

To address these gaps, we developed a computational model to study the dynamics of LAT condensation driven by TCR signaling. In this model, TCR antigen engagement leads to localized LAT phosphorylation and condensation at the cell membrane through a stochastic reaction-diffusion process. We find that this model shows all critical signaling properties of the TCR pathway, including its sensitivity to low ligand numbers, selectivity to high affinity ligands, and dynamic range in response to variations in pMHC ligand numbers. We also test key experimental predictions of this model, including activation dynamics, which provide a direct benchmark of the dynamics of pLAT cluster formation in response to single pMHC-TCR binding events[80]. Finally, we experimentally validate the spatial component of our model using a designed protein system[82] that allows systematic variation of antigen spacing independent of antigen number. Together, these findings establish a quantitative framework for understanding the exceptional functional capabilities of the T cell in threat sensing and discrimination.

Results

The LAT condensation model

To assess the contribution of LAT condensation to TCR signaling we built a dynamical model to capture the spatial distributions of LAT at the membrane due to clustering. To achieve this property, we built our model on a 2D grid representing the plasma membrane of a T-cell. Within this 2D grid, we model the location of TCRs and LAT molecules, both of which can diffuse freely in the membrane (fig. 1B). Other signaling- and condensation-related molecules such as Lck, ZAP70, SOS, Grb2, PLCγ, and CD45 are modeled implicitly by the effects they have on either TCRs or LAT molecules. In each time-step of a simulated signaling trajectory, the actions of each TCR and LAT molecule are calculated independently, based on the probabilities that each molecule performs any of the available actions. The determined probabilities are then used to stochastically make the relevant changes to the state of the simulation grid, and the next time step begins. By explicitly modeling the location of each molecule our model can capture signaling effects that occur due to spatial inhomogeneities that are not considered in existing models.

To interrogate the role of LAT condensation independently from other signaling steps, we do not explicitly model the phosphorylation steps between TCR binding and ZAP70 activation. Instead, we assume that a pMHC-bound TCR immediately phosphorylates LAT (via ZAP70) in its local neighborhood, assuming that TCR phosphorylation and subsequent ZAP70 recruitment occur rapidly compared to LAT phosphorylation (fig. 1A). This assumption is based on findings that LAT phosphorylation is a rate-limiting step for T cell signaling initiation[83,84]. In reality, upstream events do take time to occur, and are likely also involved in kinetic proofreading[85,86], especially in suppression of signals from very short binding ligands (e.g. binding dwell times less than a second) . However, because we sought to evaluate the degree to which LAT condensation can contribute to kinetic proofreading and other signaling properties, we chose to leave out the effects of other phosphorylation steps on these properties.

We assume that the pMHC-bound TCR complex phosphorylates LAT at four pY sites[87]: three sites are ideal substrates for the kinase and hence are phosphorylated at a higher rate, while one (Y132) is less ideal and is phosphorylated at a lower rate[83,88] (fig. 1A). We model the Y132 site separately from the other sites, as its slower phosphorylation has been shown to be a rate limiting step in T-cell signal initiation[83,84]. To account for these differences, we model two types of pLAT: p_4LAT, which has all pY sites phosphorylated and can participate in PLCγ-mediated bonds, and p_3LAT, which has no phosphorylation on Y132 and can bind other p_3LAT only via weaker Grb2-SOS-mediated interactions (fig. 1A,C,E). These reversible interactions are multivalent and allow LAT to form a condensate by creating a network of bonds through its pY sites[18,77,89].

pLAT condensate formation is opposed by the activity of phosphatases (such as CD45) which dephosphorylate pLAT, rendering it incapable of crosslinking (fig. 1A)[90,91]. While phosphatases act rapidly at uncondensed LAT sites, they are sterically hindered both directly by the binding of other molecules to pY sites, and indirectly through exclusion from condensates[18,17,15]. We thus modeled a slower dephosphorylation rate for pLAT that has more bonds to other pLAT, with the stronger binding of PLCγ to pY132 offering a higher degree of protection than the weaker Grb2/SOS mediated bonds.

Altogether, our model captures the key physical and chemical processes underlying TCR-induced LAT clustering, thus allowing us to assess how this emergent phenomenon contributes to the observed input/output characteristics of the TCR signaling pathway.

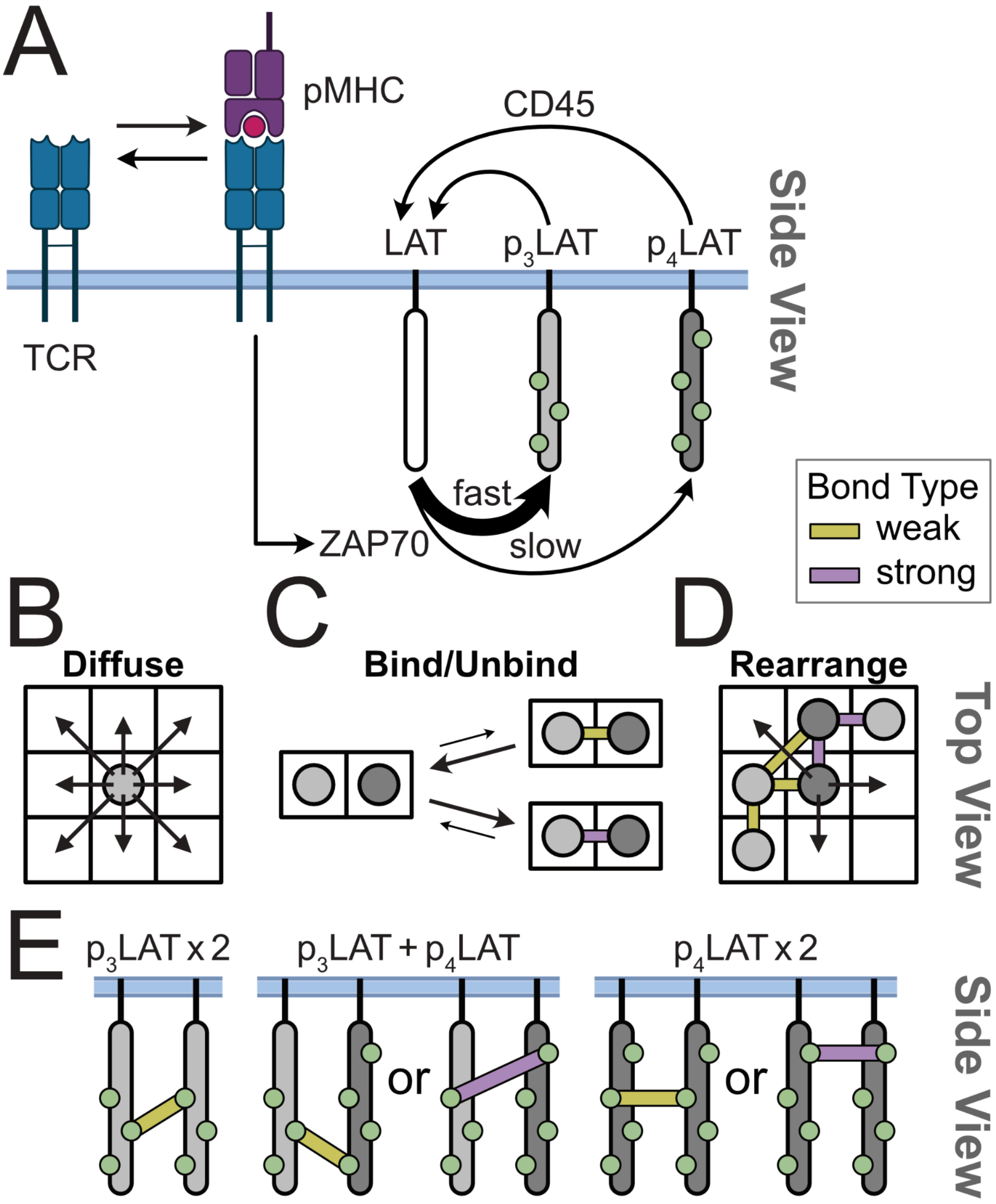

Figure 1. pLAT Condensation Model. A) The TCR module of the model consists of pMHC binding and subsequent activation of ZAP70. Active ZAP70 phosphorylates LAT to produce p_3LAT (light gray) and p_4LAT (dark gray). pLAT can be dephosphorylated by phosphatases such as CD45. B) The simulation takes place on a 2D grid over which pLAT molecules can diffuse. C) Adjacent pLAT molecules can bind to each other via cross-linking molecules that are not

explicitly modeled. Bonds can be weak (yellow bar; slow on-rate, fast off-rate) or strong (purple bar; fast on-rate, slow off-rate). D) Molecules within a pLAT cluster can rearrange internally as long as the number of bonds made between them is conserved. E) Strong bonds (purple bars) require that at least one of the bonded pLAT molecules be p_4LAT, while weak bonds (yellow bars) do not have this requirement.

Phosphorylated LAT condenses above a critical concentration

In vitro studies have shown that pLAT can undergo phase separation above a critical concentration[18]. We first ran simulations where the number of pLAT molecules was held constant by removing the ZAP70 and CD45 activity from the model to determine whether our model could recapitulate similar pLAT condensation behavior. In these simulations we expect to observe behaviors that resemble a phase transition, including the presence of a critical concentration below which clusters do not form, and above which they form rapidly[92,93]. Consistent with these expectations, we observed a sharp transition where pLAT began to form clusters above a critical concentration threshold (160/μm^2; fig. S1A,B,C; movie S1,2). Additionally, at this critical concentration, pLAT shows a very clear bimodal distribution where each simulation either achieves complete clustering or does not form clusters at all, indicating that once a condensate is nucleated it forms very quickly and is very stable (fig. S1D). Finally, we found that the mean time to first cluster decreases with concentration following a power law with an exponent of roughly 3.5, suggesting that cluster nucleation requires simultaneous binding of 3-4 pLAT molecules (fig. S1E). These results demonstrate that the model recapitulates the key aspects of nucleation and condensation as intended and allows us to further explore how these features can contribute to the behaviors of the TCR signaling pathway as a whole.

Individual bound TCRs nucleate signaling condensates with a time delay

Having shown that pLAT alone can undergo clustering in our model, we proceeded to ask whether pLAT can form clusters in response to TCR pMHC engagement and LAT phosphorylation. Single-molecule studies had previously shown that single TCRs can induce LAT clusters with a variable time delay after pMHC binding[80]. In these experiments TCRs are only tracked after they bind to a pMHC, and the peptides used are very high affinity. We therefore ran our simulations with the TCR constantly bound to pMHC throughout the simulation (fig, 2A,B) to mimic the tracked TCRs in the experiment.

From simulations, we found that single active TCRs can nucleate clusters of pLAT molecules with a time delay, consistent with experimental observations[80] (fig. 2A; movie S3). In this representative simulation (fig. 2) we found that a small cloud of pLAT molecules formed under the active TCR very shortly after the start of the simulation, with a concentration gradient that decayed with increasing distance away from TCR. While pLAT molecules in this cloud could interact with one another, these interactions were transient, such that the majority of molecules were either monomers or dimers (fig. 2B, inset). However, after a time delay, a LAT cluster became nucleated and proceeded to grow, fed by continuing TCR activity (fig. 2B). Notably, these results show that a single pMHC ligand gave rise to a large cluster containing hundreds of LAT molecules. This remarkable amplification in the number of active signaling molecules is also observed experimentally[25,80] and demonstrates the ability of a condensation-based signaling system to sensitively amplify signals from single molecule binding events.

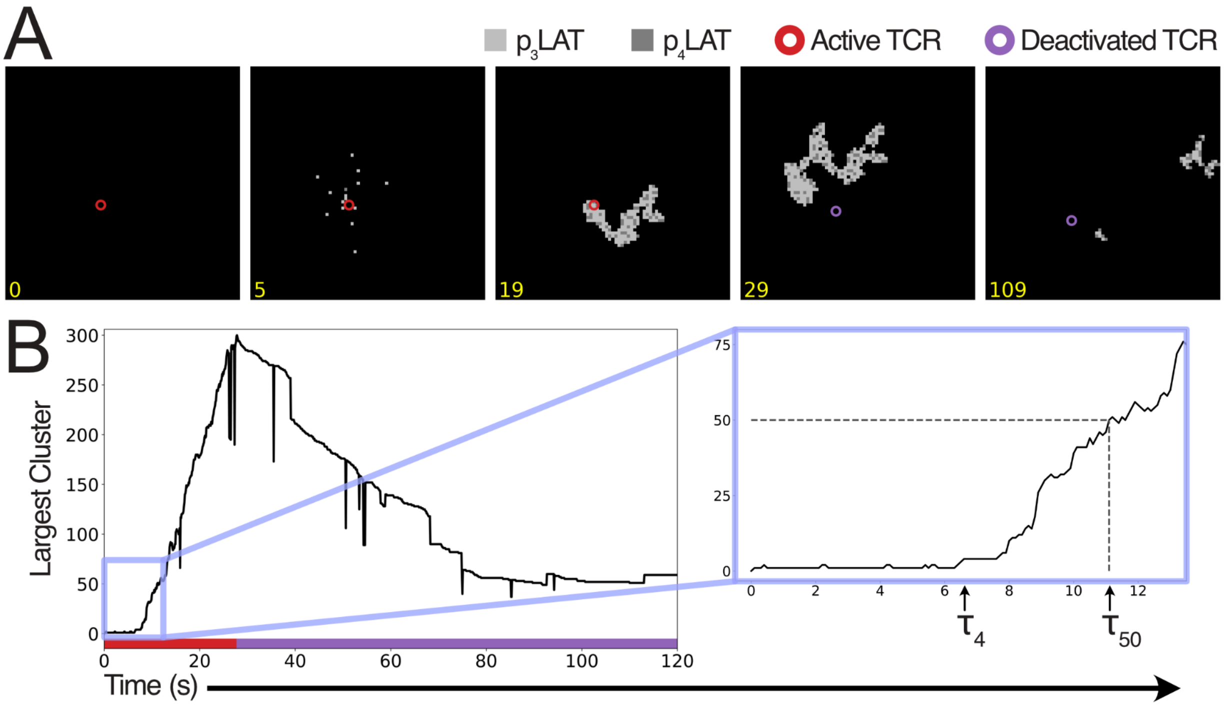

Figure 2. Individual bound TCRs nucleate signaling condensates with a time delay. A) Snapshots from an example simulation of a single permanently bound TCR. Light gray squares represent p_3LAT and dark gray represent p_4LAT. Circles represent TCRs, red when it is active, purple when it is deactivated. Each snapshot represents a 2.25μm square. Time stamps are in seconds. B) Trace of the size of the largest cluster over time in the example simulation in (A) showing the time to the first cluster of size 4 or greater (τ_4), or size 50 or greater (τ_{50}). Colored bar indicates TCR status; red for active, purple for inactive.

Simulated LAT cluster sizes and dynamics match single-molecule imaging

To further validate our model, we compared LAT clustering dynamics from simulations to those previously seen in single molecule imaging experiments[80] (fig. 3). These experiments track individual pMHC-TCR complexes and LAT clusters that form nearby; importantly, because they track the initial time that a single TCR binds to a pMHC ligand, they allow delay times and size distributions of pLAT clusters to be precisely measured. In order to compare cluster sizes between simulation and experiment, we define N_{max} to be the maximal number of pLAT molecules in clusters over the course of a simulation (fig. 3A). Our simulations show a range of N_{max} values from 50 to more than 500 molecules (fig. 3B), demonstrating the ability to amplify a single-molecule input signal into a much larger response. This range of sizes closely matches the distribution observed experimentally, where cluster sizes range from the detection limit of 200 molecules to about 700, with the majority between 250 and 400 molecules[80].

We next quantified the delay time between pMHC binding and pLAT cluster formation (τ_{50}, fig. 3C), as this time delay in signaling condensate nucleation could constitute a kinetic proofreading step in the TCR signaling pathway. Comparing the simulated and experimental distributions, we find that they are strikingly similar with an initial rise peaking around 20-25s, followed by a slow decay (fig. 3C)[80]. Critically, both distributions do not follow a monotonic decay such as the exponential distribution expected from a first-order process. This delay in LAT cluster nucleation could contribute to selectivity by filtering out short-lived binding events from weak ligands, a possibility we will test below.

Finally, in single-cell molecule experiments, pMHC-bound TCRs give rise to pLAT clusters only in a fraction of cases (25-50%), even when they were bound for very long times[80]. This cap on TCR-induced clustering likely arises from feedback mechanisms which deactivate pMHC-bound TRCs over time. As the binding lifetime increases, if a cluster has not yet formed, it becomes more likely that the TCR will be deactivated before a cluster can form. This negative feedback is incorporated into our model and sets the probability of clustering at about 27% (fig. 3D, WT), in accordance with experimental data[80].

The probability of forming a cluster, the size of the cluster that is formed, and the delay between pMHC-TCRC binding and cluster formation are all critical parameters controlling the T-cell response at a cellular and population level. The close agreement in these measurements between our simulated results and experimental data suggest that our model is able to accurately capture these key aspects of early TCR signaling.

LAT phosphorylation kinetics modulate the time delay to condensate nucleation

LAT Y132 phosphorylation has been shown to be a rate-limiting step in T-cell signaling initiation[83], and LAT mutations that accelerate this reaction have further been shown to promote LAT cluster formation[80]. Here, we test whether changes to rates of Y132 phosphorylation can modulate time delays to condensate nucleation. To do so, we simulated two mutations in LAT which are known to affect both the PLCγ binding site and the kinetics of signaling. The glycine residue at position 131 in WT LAT makes Y132 a poor substrate for ZAP70, reducing its phosphorylation rate significantly[83]. The G131D mutation increases this rate[83], which we modeled as an increase in the p_4LAT production rate in our simulations. Conversely, the Y132F mutation completely prevents phosphorylation at the same site, modeled as a reduction of the p_4LAT production rate to zero. We first observed that these mutations had the expected effect on the fraction of TCR binding events that are productive. Y132F completely abolishes clustering in simulations (fig. 3D), paralleling the dramatic reduction in signaling observed in T-cells bearing that mutation[87,94]. Conversely, G131D increases the productive binding rate from 27% to about 75% (fig. 3D), in agreement with the observation that T-cells bearing this mutation are more easily activated[83].

In addition to affecting the probability of clustering, G131D dramatically reduces nucleation delay times. Both simulated and experimental results show an earlier peak in the delay time distribution around 15s and a shorter tail to the right of the distribution (fig. 3E)[80]. The shortened delay suggests a decreased ability for G131D cells to discriminate between weak and strong binding pMHCs.

Through these simulations we show that our model recapitulates the effects of mutations in LAT that affect the kinetics of its phosphorylation at the key Y132 site. We are able to recapitulate expected effects on the frequency of cluster formation as well as on the distribution of cluster initiation times. And more broadly, we are able to show that the kinetics of LAT phosphorylation have a direct influence on the signaling outcome, and are likely to influence kinetic proofreading via modulation of delay time distributions.

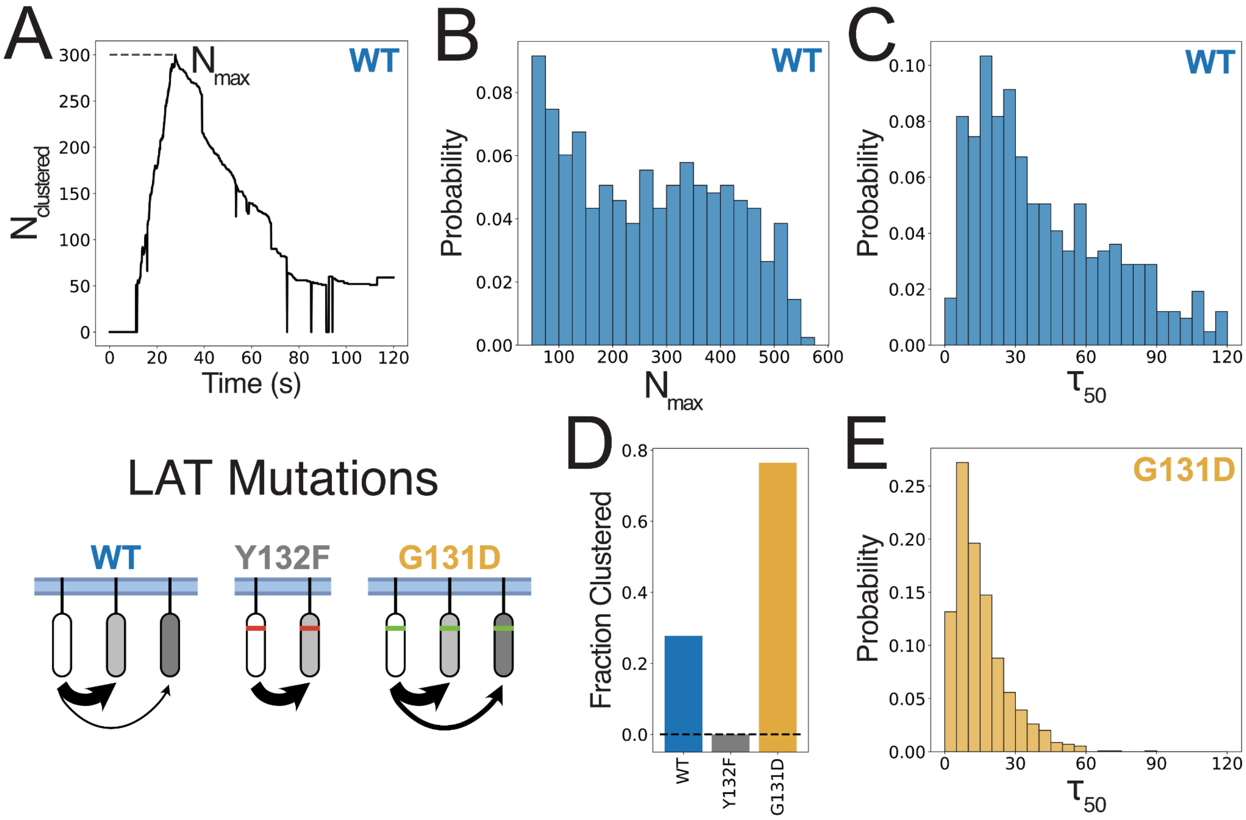

Figure 3. Simulated LAT clustering dynamics match experimental results. Data from simulations of single permanently bound TCRs on cells with varying LAT mutations. 1500 replicate simulations were run for each mutant. A) Representative trace of the number of pLAT molecules in clusters of size 50 or greater in a single simulation with WT LAT. N_{max} denotes the maximum number of clustered molecules over the course of the simulation. B) Distribution of N_{max} values over replicate simulations with WT LAT. C,E) Distribution of τ_{50} values (the time it takes to form a cluster of size 50 or greater) in simulations for either WT LAT (C) or G131D LAT (E). Simulations in which no cluster forms are omitted from these distributions. D) Bar chart showing the fraction of simulations that produce a LAT cluster for various LAT mutants.

Kinetic proofreading by condensate nucleation in TCR signaling

The observed delay between pMHC binding and LAT condensation can potentially serve as a kinetic proofreading step and thereby enable T-cells to selectively respond to high-affinity pMHC ligands. To test this idea, we simulated a system with diffusible TCRs on a 2D membrane that were then stimulated with pMHCs with varying lifetimes of TCR binding. In order to allow a fair comparison between different affinity pMHCs, we increased the simulated concentration of pMHC inversely with its TCR binding lifetime, thus keeping TCR occupancy the same (fig. 4A,C). In these simulations, the fraction of pMHC-bound TCRs remained the same as binding lifetimes increased, as expected (fig. 4C). In contrast, LAT clustering occurred preferentially in response to stimulation with tight-binding pMHCs, with little or no clustering occuring for pMHC ligands with shorter binding lifetimes ($\tau_{off} < 1s$) (fig. B, D-E; movie S4-10). Furthermore, as pMHC binding lifetimes increased, both the likelihood of cluster formation as well as cluster sizes increased, indicating that the system is able to encode pMHC affinity information at this step. These results directly demonstrate kinetic proofreading by LAT condensation. Because the TCR occupancies are matched across all simulations, the reduced clustering observed with the faster off-rates can only be due to a kinetic proofreading mechanism. To the extent that this occurs in these simulations, the proofreading is carried out solely by LAT condensation, and not any upstream components of the pathway since they are not considered in our model. These results show that LAT condensate nucleation is sufficient to generate the time delays needed to enable selective response to only high-affinity pMHC ligands.

To probe the role of slow LAT phosphorylation in setting time delays in kinetic proofreading, we ran the same TCR occupancy-matched simulations with the G131D and Y132F mutants and looked at the effects of these mutations on signaling magnitude and proofreading. As expected, the G131D mutation increased the LAT clustering response both in terms of the fraction of simulations that were able to form clusters (fig. 4E) and the size of those clusters (fig. 4F), while the Y132F again completely prevented clustering. Strikingly, the G131D mutation drastically reduces the ability of the simulations to distinguish self-peptides (τ_{off}<1s) from foreign peptides (τ_{off}>10s) with roughly 20% of G131D simulations responding to peptides with 50ms half-lives (fig. 4E). The reduced degree of proofreading caused by this mutation explains why mice bearing this mutation show enhanced responses to weak pMHC stimuli and develop autoimmune disease[30].

Together, these results indicate that LAT nucleation constitutes a key kinetic proofreading step in T cell signaling . The delay imposed by the rarity of nucleation events provides a simple mechanism for filtering out short-lived pMHC-TCR interactions. Additionally, the phosphorylation rate of the PLCγ binding site is a critical parameter that controls the threshold for binding lifetimes that are able to produce a condensation response.

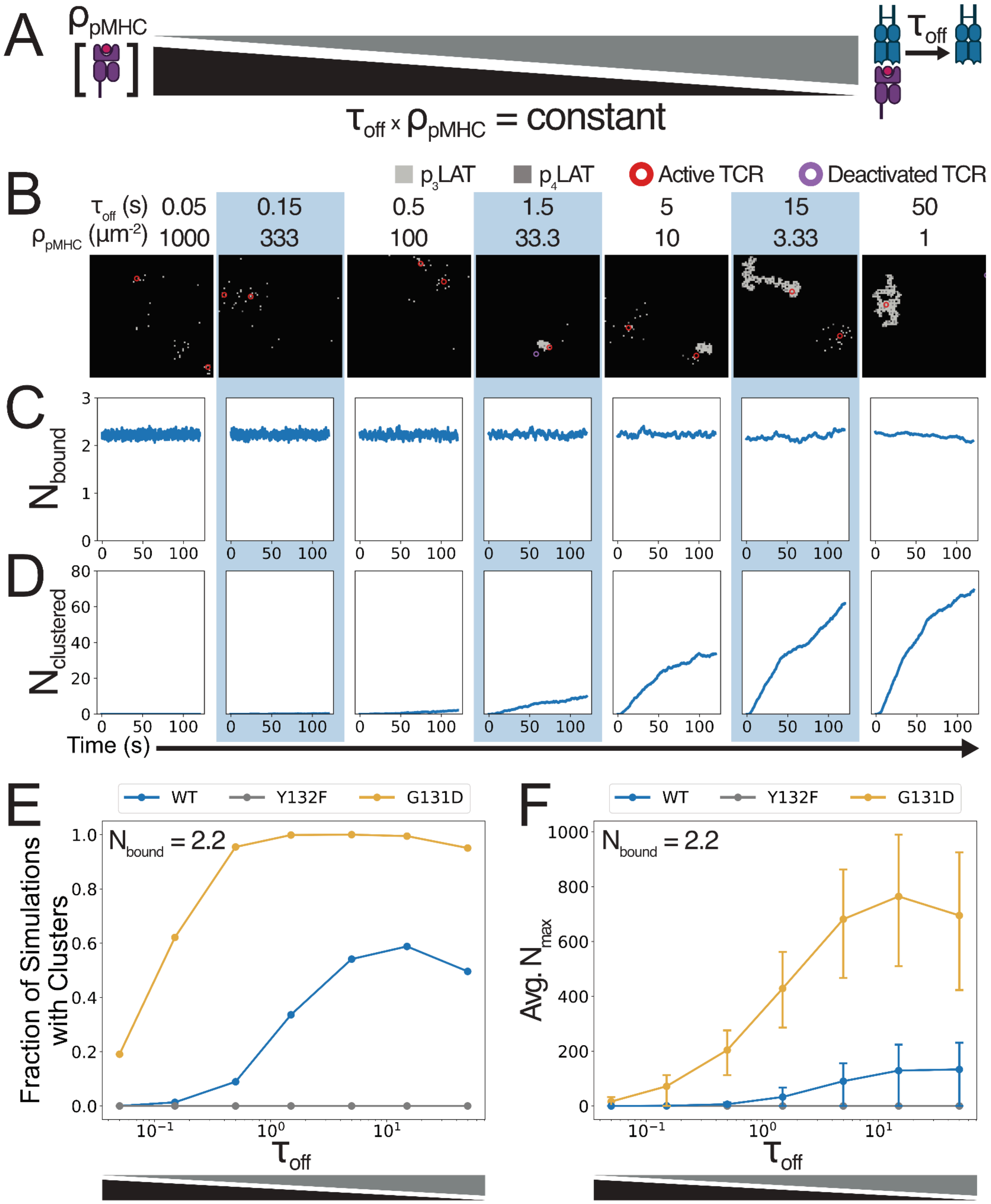

Figure 4. Kinetic proofreading by condensate nucleation in TCR signaling. Data from simulations for varying LAT mutants and pMHC lifetimes, with TCR occupancy held constant. 750 replicate simulations were run for each condition. A) Schematic showing the simultaneous increase in pMHC-TCR bond lifetime and decrease of pMHC 2D density used to achieve constant numbers of bound TCRs. B) Snapshots of representative simulations at each pMHC

off-rate with WT LAT. Light gray squares represent p_3LAT and dark gray represent p_4LAT. Circles represent bound TCRs, red for active and purple for deactivated. Each snapshot represents a 2.25μm square. C) Average number of bound TCRs over time for simulations at each pMHC off-rate with WT LAT. D) Average number of pLAT molecules in clusters of size 50 or greater for the same simulations in (C). E) The fraction of simulations for each condition that produce a cluster of size 50 or greater. F) Mean N_{max} over replicate simulations for each of the simulated conditions. Error bars show 25th and 75th percentiles of replicate simulations.

Cluster growth allows wide dynamic range

In addition to distinguishing between ligands of varying affinity, T-cells also need to respond to variations in pMHC levels[75], which may indicate the severity of the infection. Most current models propose that while signaling is digital at the single cell level, the population can have an analog response where the pMHC dose tunes the percentage of cells that respond[20]. However, some aspects of TCR signaling have been demonstrated to respond to pMHC dose in a graded manner at the single cell level[75,95].

To determine the extent to which LAT clustering can convey information about pMHC dose, we tested the response of our model to antigen doses spanning three orders of magnitude (fig. 5). We first found that each simulated cell has a binary response in terms of the presence or absence of a cluster, with the fraction of responding cells increasing as with the antigen level and lifetime (fig. 5A). At low doses, the average size of clusters, when they are present, remains low, but as the fraction of cells responding begins to saturate, the size of the cluster begins to grow (fig. 5B). These results indicate that our model is able to produce both all-or-none signaling responses in the nucleation of a LAT cluster, as well as graded increases in the sizes of nucleated LAT clusters with increasing antigen dose. Together, these results explain how LAT clustering dynamics could give rise to digital activation for signal amplification and single molecule sensitivity, as well as analog tuning for dynamic range in pMHC dose detection.

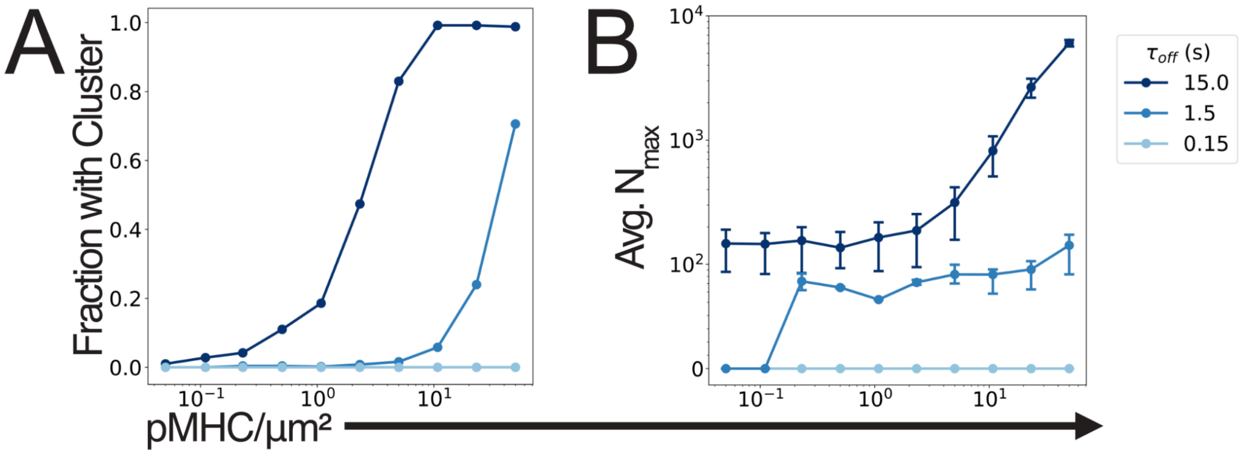

Figure 5. pLAT condensation increases gradually with increasing pMHC stimulation. Data from simulations with varying pMHC dose and lifetime, with WT LAT. 500 replicate simulations were run per condition. A) The fraction of simulations that produce pLAT clusters in response to

stimulation with varied pMHC density at each of three TCR-pMHC half-lives. B) The mean N_{max} in simulations where a cluster forms for the same stimulation conditions as (A). Error bars show 25th and 75th percentiles of replicate simulations.

LAT condensation sets an optimal pMHC spacing for T cell activation

The analysis above shows that our LAT condensate nucleation model can uniquely give rise to sensitivity, specificity and dynamic range in TCR signaling. This model agrees with existing experimental data; however, we have not yet shown that there are unique predictions made by this model that can allow it to be distinguished from other existing kinetic proofreading models. Most other dynamic models make the assumption of a well-mixed system, where a scalar value can represent the concentration of each signaling molecule. In these models, the number of long-lived binding events is the only relevant parameter for activation[21,22,20,23,25], and parameters such as the spacing between antigens would not affect signaling activity, despite evidence that this may be important[96–98].

To test if our model made any unique predictions about spatial dependencies in TCR signaling, we ran simulations where a square grid containing a fixed number of pMHCs was placed opposing the T-cell so that TCRs could only bind and activate in those locations (fig. 6A). We varied the 2D density of pMHCs over roughly three orders of magnitude by adjusting the spacing between them, while keeping the number of pMHCs constant, and examined the effects on LAT clustering (fig. 6B-C). In contrast to other models, our simulations show that there is an intermediate spacing of pMHCs that is optimal for LAT cluster formation and T-cell signaling (fig. 6C). At both low and high pMHC density the cluster sizes are smaller, while only intermediate densities achieve maximal clustering (fig. 6C; movie S11-15). At very low pMHC densities, each pMHC-TCR complex is isolated from others such that pLAT produced at one TCR is dephosphorylated before it can reach a cluster formed at another TCR (fig. 6D. As the pMHC density increases, TCRs are able to cooperate, adding pLAT to clusters that were nucleated by other TCRs, allowing the cluster to grow larger than what could be produced by a single pMHC-TCR complex (fig. 6E). However, if the pMHCs are brought too close together, then the TCRs become crowded, merging into what is effectively a single TCR that produces pLAT at a higher rate. This allows a cluster to form faster, but the cluster size is still limited to what can be produced by a single TCR (fig. 6F). Thus, LAT condensation sets the spatial scales of TCR activation in addition to the temporal scales discussed earlier.

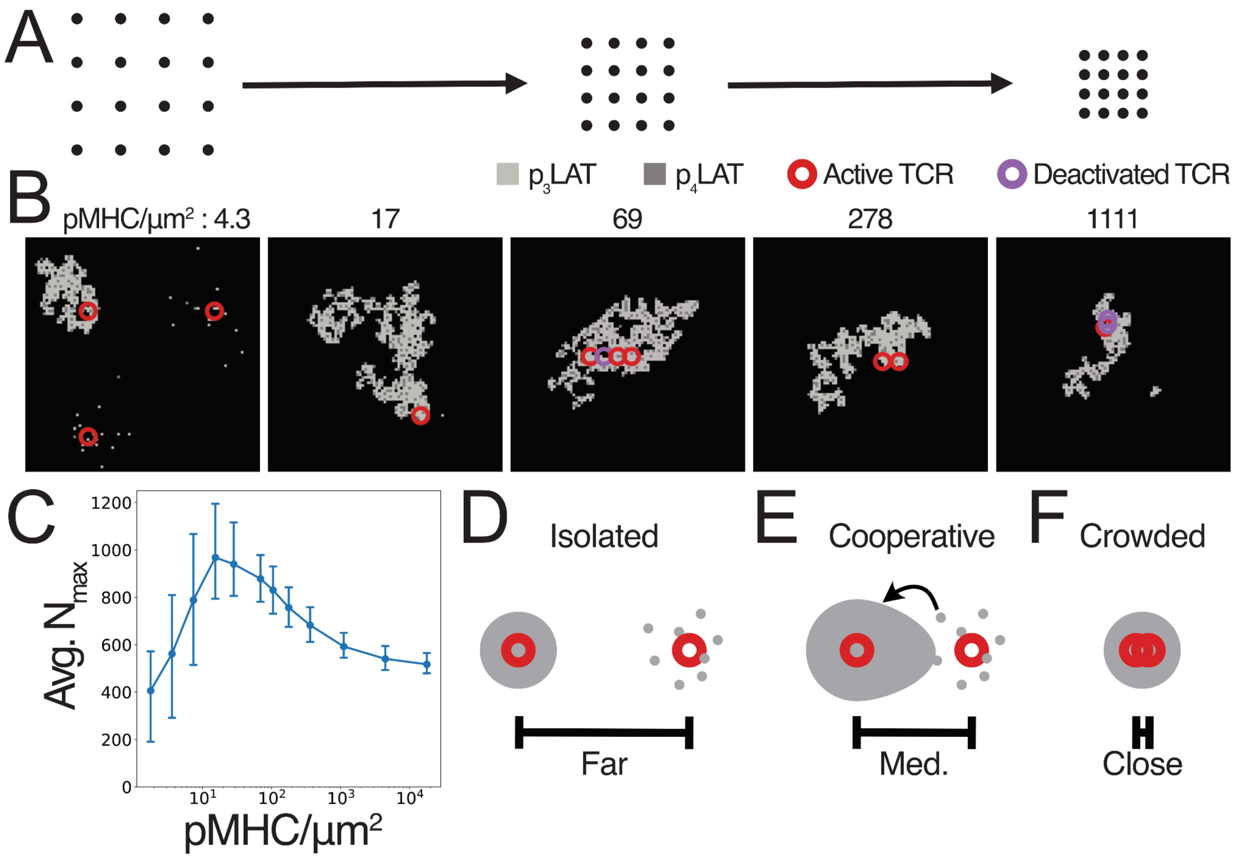

Figure 6. LAT condensation simulations predict an optimal pMHC spacing for TCR activation. Data from simulations with 16 fixed pMHCs at varying spacing, with WT LAT. 500 replicate simulations were run for each pMHC density. A) Schematic of pMHC placement in simulations of varying antigen density at constant antigen number. B) Snapshots from example simulations at different pMHC densities. Light gray squares represent p_3LAT and dark gray represent p_4LAT. Circles represent bound TCRs, red for active and purple for deactivated. Each snapshot represents a 1.27µm square. C) Mean N_{max} values for each pMHC density. Error bars show 25th and 75th percentiles of replicate simulations. D-F) Cartoon representation of the underlying cause of optimal spacing. Large gray circles represent pLAT clusters, small gray circles represent individual pLAT molecules, and red circles represent TCRs. D) At low pMHC density, pMHC-bound TCRs are isolated, preventing pLAT produced at one TCR from augmenting a condensate at the neighboring TCR. E) At intermediate pMHC density pLAT produced at one pMHC-bound TCR can add to a condensate nucleated at a neighboring TCR. F) At high pMHC density, neighboring pMHC-bound TCRs are crowded, limiting the area over which a condensate can form.

2D protein arrays allow precise tuning of pMHC spacing independent of dose

In order to test the prediction of optimal pMHC density, we developed an imaging assay to probe the relationship between antigen density and LAT clustering. As a platform for antigen presentation we used a 2D protein array[82] to place pMHCs at regular intervals (fig. 7A). The arrays consist of two components in addition to the pMHC: the A component which is fused to the SpyCatcher protein[99], and the B component which is fused to GFP. When mixed, the A and B components self-assemble into a hexagonal grid (fig. 7A). By controlling the stoichiometry of the SpyTag-pMHC and SpyCatcher-A prior to assembly with component B, we can control the 2D density of pMHCs on the arrays. Furthermore, by quantifying stochastic variations in the sizes of arrays we are able to assess effects of variations in pMHC number, thus allowing us to independently assess the effects of pMHC spacing and number on TCR signaling.

As a model antigen, we presented the gp33-41 peptide from Lymphocytic Choriomeningitis Virus (LCMV) complexed to the murine class I MHC, H-2D(b)[100]. We used as a model a Jurkat T-cell line expressing the cognate P14 transgenic TCR and the CD8 co-receptor, which we found gives rise to robust downstream signaling[75]. After the arrays were assembled, we seeded them onto a glass imaging surface and plated the P14 Jurkat cells on top of them. After a brief stimulation, we fixed the cells and stained them for phospho-CD3ζ (pCD3ζ) and phospho-LAT (pLAT; fig. 7B,C). Each array produced a bright GFP spot in the image which we used to locate the regions of the image where arrays overlapped cells. We additionally calibrated our measurements of GFP brightness in order to be able to count the number of GFP molecules and, by extension, the number of pMHCs present in each array (fig. S2)[101]. Using these values, along with several other measurements, we filtered the arrays to remove aggregates and imaging noise, yielding a set of high quality arrays that we used in all downstream analysis (fig. S3).

We first used the median phosphorylation signal in each array as a measure of the effectiveness of array stimulation (fig. 7D,E). These measurements showed a significant elevation of pCD3ζ and pLAT levels in arrays that have a high density of pMHC on them, while those with no pMHC did not induce any phosphorylation above background (fig. 7D,E). Additionally, larger arrays (higher GFP copy number) induced higher levels of phosphorylation (fig. 7E). These results confirm that we were able to stimulate Jurkats with the arrays in a pMHC-dependent manner, and that phosphorylation is dependent on the number of pMHCs that contacted the cell. We note that array-displayed pMHCs differ from those displayed on cell membranes in their TCR contact heights, as well as their ability to exclude membrane-associated proteins such as phosphatases. These kinds of differences have been shown to impact TCR signaling initiation[97,102] and could affect the fit of the model to the resulting data because the model was adjusted to match data from membrane-bound pMHC stimulation. We therefore compare only general trends in these experiments with the model predictions instead of precise values.

Consistent with the differing biochemical mechanisms for activation of CD3ζ and LAT, the distributions of their phosphorylation levels show different shapes in response to pMHC stimulation (fig. 7E). The pCD3ζ distribution shows a clear increase in its mode relative to the background distribution as the array size increases, consistent with CD3ζ phosphorylation occurring in proportion to the number of pMHCs present. The pLAT distribution, on the other hand, shows a widening tail to the right without a major shift of the mode. This distribution is consistent with pLAT signals resulting from stochastic cluster nucleation in a fraction of cases that increases with pMHC number. These results are consistent with our model prediction that

LAT signaling condensates nucleate only for a fraction of pMHCs (fig. 3C), but can grow to large sizes that scale with the number of initiating TCR inputs (fig. 5).

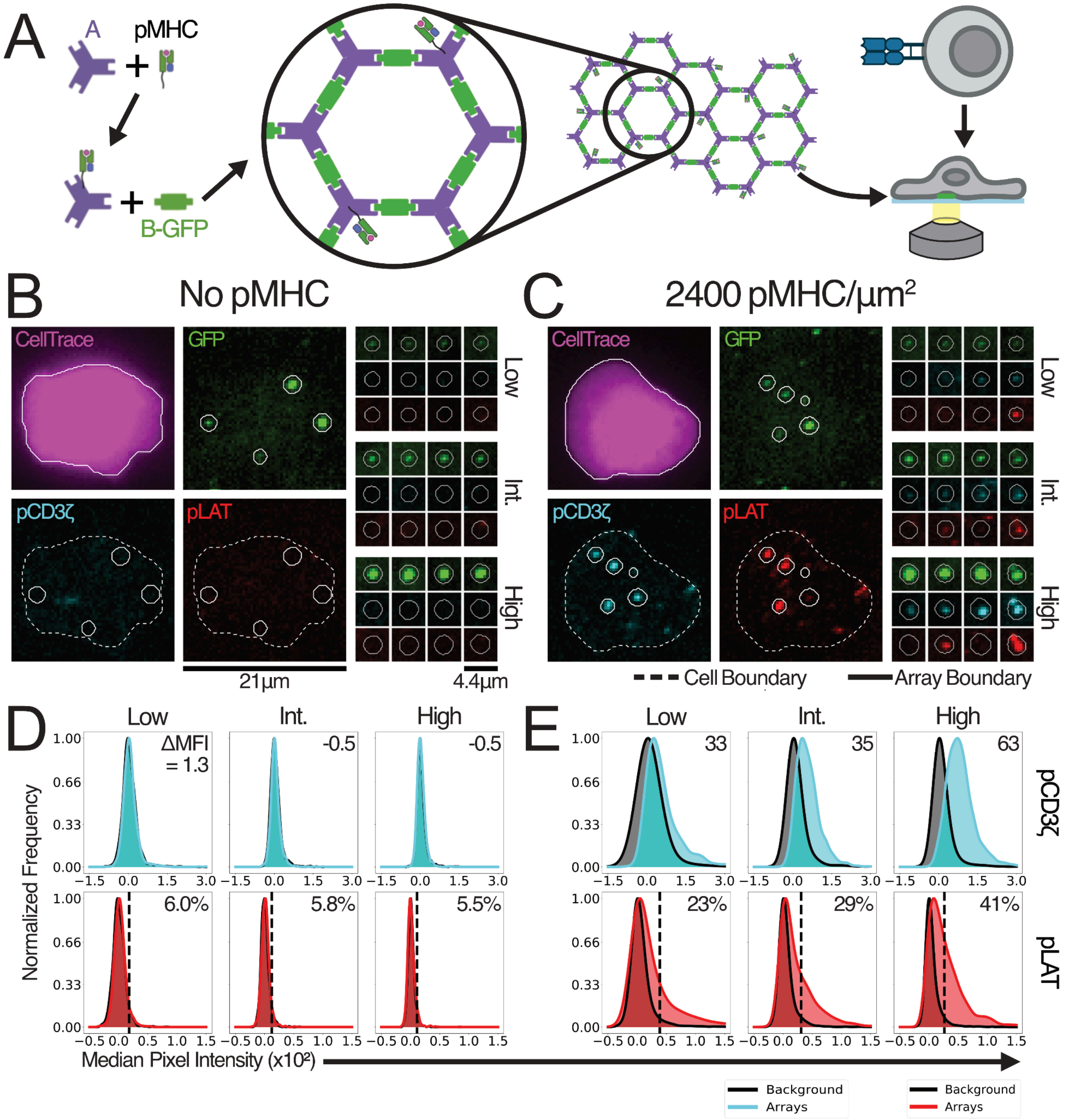

Figure 7. 2D protein arrays enable precise pMHC display. A) Illustration of the array assembly and imagining protocol. Components A and B are shown as purple triangles and green rectangles, respectively. The assembled array has a 31nm distance between opposing edges of each hexagonal unit[82]. B,C) Example images of cells (purple) and arrays (green) that were fixed 5min after interacting with blank (B) or pMHC-presenting arrays (C) showing the pCD3ζ (light blue) or pLAT (red) responses to pMHC array stimulation. Larger fields of view (left) show a single representative cell interacting with arrays. Smaller fields of view (right) show four representative arrays with either low (33rd percentile), intermediate (int.; 66th percentile), or high

(100th percentile) GFP intensities. D,E) Normalized kernel density estimates for the distributions of median pCD3ζ (light blue, top) or pLAT (red, bottom) signal for arrays with low (left; 0-33rd percentile) intermediate (middle; 33rd-66th percentile) or high (right; 66th-100th percentile) GFP intensities, compared to the same measurement in random locations (gray) for blank (D) or pMHC-presenting arrays (E).

pMHC arrays reveal optimal antigen spacing for LAT condensation

With the functionality of the 2D pMHC arrays validated, we moved forward to test whether varying pMHC density has an impact on downstream signaling. To do this we created arrays with pMHC density varying from 37.5 to 2400 pMHC/μm^2 and measured the CD3ζ and LAT phosphorylation responses to each density. We first examined the relationship between array size and phosphorylation signal and found that, for all pMHC densities above zero, phosphorylation increased with increasing array size (fig. S4A). The slope of this relationship also increased with increasing pMHC density (fig. S4B). These results show increasing pMHC number results in a stronger downstream CD3ζ and LAT phosphorylation; more antigen results in a stronger signal.

After demonstrating that pMHC array stimulation produces the expected response to varying pMHC number, we next tested whether TCR signaling varies with pMHC spacing at fixed pMHC number. To compare arrays across all pMHC densities despite non-overlapping distributions of pMHC number when comparing the lowest densities to the highest, we measured the activation efficiency for arrays of each density. This metric is defined as the slope of the line of best fit to the relationship between phosphorylation intensity and pMHC number (fig. S5), and measures the degree to which pMHCs elicit signaling, at a given spacing. We observed that pMHC spacing does not have a strong effect on CD3ζ activation efficiency (fig. 8A), while pLAT activation efficiency shows a clear peak at intermediate pMHC spacing (fig. 8B). This behavior is very similar to the behavior predicted by our model (fig. 6). The phosphorylation of CD3ζ, which does not participate in LAT condensation, is not impacted by pMHC density. Conversely, LAT phosphorylation is dependent on pMHC density, and requires an optimal density for maximal signaling. These experimental and simulated results further demonstrate that LAT condensation is central to the TCR signaling pathway. By examining the experimental and simulated responses to varied antigen density we have demonstrated that the optimal antigen spacing is controlled by the underlying length scales of LAT clustering. Thus, LAT condensation controls both the length and time scales required for optimal T-cell activation.

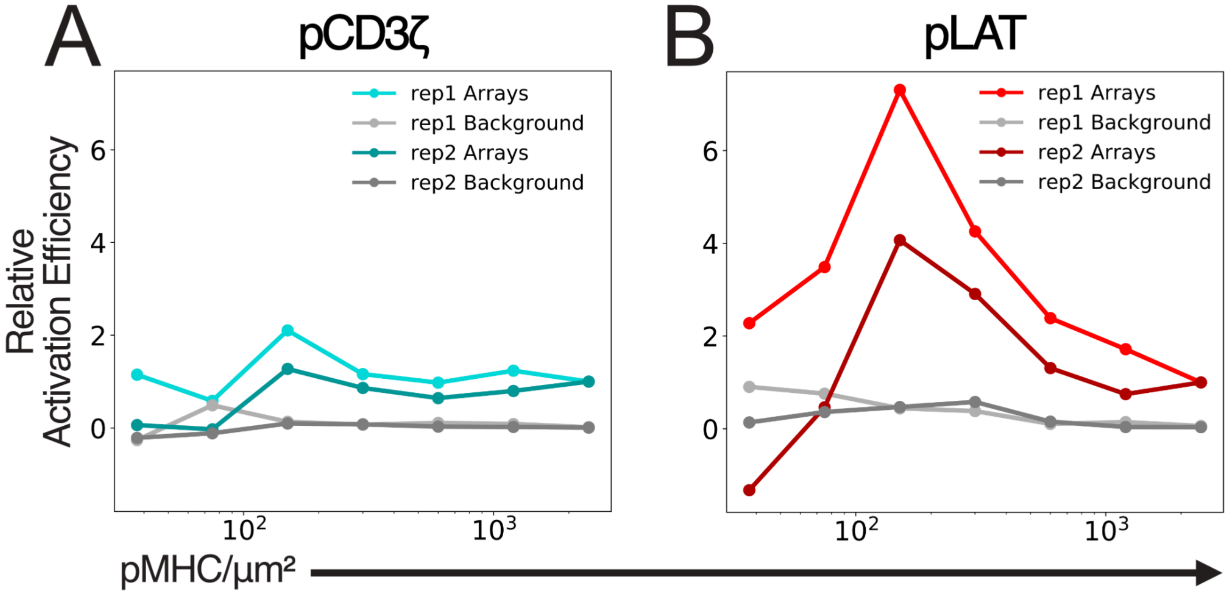

Figure 8. 2D arrays reveal optimal pMHC spacing for LAT phosphorylation. A,B) Line plots showing CD3ζ (A) or LAT (B) activation efficiency as a function of pMHC density for arrays (color) or background (gray). Activation efficiency is defined as the slope of the line of best fit for the relationship between integrated phosphorylation intensity and pMHC number. Data from two biological replicates (rep1 and rep2) are shown, with each replicate independently normalized to the efficiency at 2400pMHC/µm^2.

Discussion

The results presented here demonstrate that condensation can explain the key features of the TCR signaling network, simultaneously achieving sensitivity, selectivity and dynamic range. Prior work in modeling these properties has generally focused on the selectivity of the pathway[20–23,25,29] with some studies also giving attention to the sensitivity[20,23,26] and very few considering dynamic range[21]. The focus on selectivity likely stems from the key role it plays in preventing autoimmunity, and the fact that it is difficult to balance all three of these properties in a single model. Each property imposes its own set of constraints on the underlying signaling network, and these constraints are often in opposition to each other. Large numbers of proofreading steps have been shown to increase the selectivity of the pathway[22,25], but recent work has demonstrated that these additional steps make the system much more noisy, and can actually decrease the ability of the pathway to distinguish between weak and strong binding amid the stochasticity of biochemical reactions[29]. Strong positive feedback can significantly improve the sensitivity of a signaling pathway, but can create a digital system that is either “on” or “off” and lacks dynamic range. Strong positive feedback is also more likely to amplify small, spurious responses to off-target signals, reducing specificity. Conversely, negative feedback can improve selectivity and can explain the observation that antigens with intermediate affinity can antagonize the signaling of higher affinity antigens[20,23]. The ability of these additional feedback loops to strengthen our model’s proofreading and improve its performance could be explored in future work.

Our model is able to balance the constraints imposed by each property of the TCR signaling pathway in a single signaling step by taking advantage of the inherent physical properties of condensation. Simulations of single pMHC-TCR interactions show how condensation itself provides a strong positive feedback mechanism that can amplify even a single pMHC binding event into a large LAT cluster. The resulting cluster is able to grow in proportion to the amount of available pLAT, giving the system a wide dynamic range. We have also demonstrated that the positive feedback inherent in condensation is carefully controlled by another condensation-related mechanism: nucleation delays. Because of strong ambient phosphatase activity, pLAT is very short-lived outside of clusters resulting in fast dissipation of any pLAT that accumulates during short pMHC binding events, preventing spurious activation and filtering out noise. However, if a pMHC can bind long enough that a LAT cluster can nucleate, that cluster can grow quickly and survive even after the pMHC dissociates, allowing time for downstream signaling to be fully activated. In addition to these well-recognized properties of the TCR pathway, we have demonstrated that the same condensation mechanism underlies the spatial preferences of TCR signaling. The area over which a single TCR can produce a LAT cluster sets the optimal distance between active TCRs. The preference for intermediate antigen spacings that we observed in our model and imaging experiments is also consistent with previous work which used DNA origami to cluster defined numbers of pMHCs together[98]. The DNA origami results showed that clusters with intermediate pMHC numbers are most efficient for downstream activation[98].

It is striking that a single step in this complex signaling pathway can achieve such a wide variety of signaling properties in contrast to other models that require a complex multi-step network of reactions to achieve the same results[21,22,20,23,25]. Our model provides an intuitive explanation for the properties of TCR signaling that make T-cells such an effective component of the immune system. It also provides insight into how the system is carefully tuned to its purpose, and how small adjustments to key parameters can lead to dysfunction. The insights provided here not only paint a more complete picture of the TCR signaling mechanism, but also provide avenues to adjust T-cell signaling in therapeutic contexts. By carefully controlling LAT clustering, T-cell signaling could be modulated to suit the situation. In this way, our model can be used to both further the understanding of TCR signaling and provide guidance on how to best engineer it to fill therapeutic needs.

Methods

Computational model

Our computational model is based on a 2D square grid representing the inner leaflet of the plasma membrane. We chose to use a grid spacing of 30nm to approximately match the space a single molecule of LAT occupies in the densest LAT clusters observed experimentally. Within this grid, TCRs and pLAT molecules are modeled as occupying a single grid square to the exclusion of molecules of the same type. All other molecules are not present in the model, and their effects are modeled implicitly in the behavior of the TCRs and pLAT molecules. Each timestep of the model proceeds through the following steps:

1. TCR diffusion
2. pLAT dephosphorylation (CD45 activity)
3. TCR-pMHC binding and activation
4. LAT phosphorylation (ZAP70 activity)
5. pLAT diffusion
6. Intra-cluster rearrangement
7. pLAT-pLAT binding (via Grb2, SOS, and PLCγ)

These steps are repeated for the desired number of time steps (for a total of 2-5min of simulated time). The length of the time step is set to 125μs so that the probability that any molecule would take more than one action within each step of the simulation is negligible. Further details of each step are outlined in the modeling appendix of the supplemental information.

Array component purification

SpyCatcher-fused component A (A-SC) and GFP-fused cyclic component B (B-c-GFP) were expressed in *E. coli* as previously described[82]. *E. coli* cell pellets were resuspended in TBS supplemented with 5% (w/v) glycerol (TBSG), 1mM PMSF, and 0.7U/mL RNase A and lysed by sonication. Lysate was clarified by centrifugation at 14000xg for 30min and His-tagged array components were purified form the soluble fraction by Ni affinity chromatography in TBSG with either 40mM imidazole (wash) or 500mM imidazole (elution). Eluted components were further purified by size exclusion chromatography on a Superose 6 Increase 10/300 G/L column (Cytiva 29091596) to select the components that assembled into the correct oligomeric state.

Array assembly

Refolded SpyTag-fused H-2D(b) bound to the LCMV gp33 peptide (KAVYNFATM) (ST-pMHC) was provided by the NIH Tetramer Facility. Varying molar ratios of ST-pMHC and purified A-SC from 1:3 to 1:192 (ST-pMHC:A-SC) were mixed and incubated overnight at 4C to allow formation of the SpyCatcher-SpyTag bond. The resulting A-pMHC and an unreacted A-SC control were then each mixed with B-c-GFP and diluted to a final concentration of 5uM A-SC and 2.5uM B-c-GFP in TBSG with 500mM imidazole. The final mixture was allowed to assemble overnight at 4C, protected from light. Assembled arrays were separated from free components by centrifugation at 5000xg for 5min. After aspirating the supernatant, the pelleted arrays were resuspended in TBSG with 500mM imidazole to their original volumes before aspirating.

Cell culture

P14 TCR transgenic Jurkat cells were provided by Matthew Wither and grown in RPMI (Gibco 11835-030) supplemented with 10% FBS (VWR 89510-186), 10mM HEPES (Gibco 15630-080), 1mM sodium pyruvate (Gibco 11360-070), and 1x Pen/Strep/Glu (Gibco 10378-016) (T-cell media). Cells were stored as 500uL aliquots at 5M/mL in T-cell media with 5% DMSO in a liquid nitrogen dewar. Four days prior to imaging, cells were thawed, washed with 2.5mL of T-cell media, plated in 5mL T-cell media in a 6cm dish and incubated for two days at 37C and 5% CO_2. After two days cells were counted and diluted to 100k cells/mL in T-cell media and incubated for another two days at 37C, 5% CO_2.

Jurkat stimulation and staining

All stimulation and staining was performed in a 96-well black-sided glass-bottom plate (MatTek PBK96G-1.5-5-F). The plate was first coated with a mixture of anti-GFP (100x dilution; MBL Life Science 598), anti-LFA1 (10ng/uL; BioLegend 301202), and retronectin (20ng/uL; Takara T100B) in PBS overnight at 4C. After coating, the plate was blocked in 1% casein (Sigma Aldrich C8654) for 1hr at 37C and washed three times with PBS. Each well of the plate was then coated with 50uL of a 2000x dilution of assembled arrays in TBSG with 500mM imidazole. The plate was centrifuged at 100xg for 5min to improve array coating and washed three times with PBS with 0.1% CHAPS, and three times with PBS to remove any arrays and unassembled components not firmly attached. P14 Jurkats were seeded into the plate at 35k cells/well and centrifuged at 100xg for 1min to induce contact with the arrays. After 5 or 10min of incubation at room temperature (including the 1min centrifugation) cells were fixed with BD fixation buffer (BD Biosciences 51-2090KZ) for 30min at room temperature. Fixed cells were permeabilized and blocked with a 1x BD wash/perm buffer (BD Biosciences 51-2091KZ; diluted from 10x in Fc blocking buffer) for 15min at room temperature, protected from light. Cells were then washed twice with 1x BD wash/perm buffer diluted in PBS, and stained overnight at 4C, protected from light, with 50uL/well of PE-anti-pLAT (BD Biosciences 558487; 5x dilution) and AF647-anti-pCD3z (Abcam ab237452; 100x dilution) in 1x BD wash/perm buffer diluted with PBS. Staining solution was removed with two washes with BD perm/wash diluted in PBS and one with PBS. Cells were then stained with 100uL of 5µM CellTrace Violet (Invitrogen C34557) in PBS for 20min at room temperature, protected from light. Excess CellTrace was removed with three washes in PBS, and cells were left in PBS for imaging. All washes and buffer changes were performed with 100uL per well unless otherwise listed and all liquids were added gently to avoid dislodging the cells from the glass surface.

Imaging

Images of fixed and stained cells were acquired on a Leica DMi8 with a spinning disc confocal, hardware autofocus, and a 63x (0.75 NA) glycerol objective. At each position, a z-stack of 6 slices spaced out by 200nm, centered on the glass surface was acquired. Images of cells (DIC, 100ms exposure), CellTrace (405 Ex, 440/40 Em, 500ms exposure), arrays/GFP (470 Ex; 510/50 Em, 900ms exposure), PE-anti-pLAT (555 Ex, 600/50 Em, 200ms exposure), and AF647-anti-pCD3ζ (640 Ex, 700/75 Em, 900ms exposure) were acquired at each slice using an LDI-7 from 89North at 50% power.

Image Processing

Images were processed using a custom MATLAB script to extract information about the CD3ζ and LAT phosphorylation in areas where cells came in contact with pMHC arrays. First, a segmentation of the cells was determined from the CellTrace stain, including dimmer regions containing thin regions of cytoplasm in contact with the glass surface. Next, a segmentation of the arrays as determined from the GFP signal in a manner that allows detection of both very bright and very dim arrays. Finally, the total or median fluorescence intensity in each channel (with the local background in a small ring around the array subtracted from each pixel) was calculated for each region where the two segmentations overlapped. As an image-internal control, the shape of each overlapping region was extracted and placed at a random location within a cell in the same image and the same total fluorescence quantification was made in that region.

GFP Counting

To estimate the number of GFP molecules present in an array from its total intensity, we used particles with known GFP copy number to calibrate our imaging protocol. Nanocages with either 60 or 120 GFP copies attached[101] were coated onto the glass imaging surface and fixed using the same protocol as the arrays. GFP nanocage images were segmented and quantified using the same custom scripts as the array images and the average total intensity for all segmented nanocages was calculated for each GFP copy number. A linear fit to these data allows interconversion between total GFP fluorescence intensity and GFP copy number (fig. S2).

Supplemental Information

Modeling Appendix

1. TCR Diffusion

TCRs diffuse across the simulation grid from one square to an unoccupied adjacent or diagonal square with a probability, $P_{dif,f}$ defined below:

$$P_{diff} = D\frac{\Delta t}{(\Delta x)^2} \tag{1}$$

Where Δt is the simulation time step and Δx is the width of the grid square, and D is the diffusion coefficient of the TCR. Based on experimental measurements, we set the diffusion coefficient of pMHC-bound TCRs to be roughly 15 times slower than free TCRs[103].

Table S1. TCR diffusion parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability. When a parameter was set based on a prior experimental measurement, a citation is listed for that measurement.

Parameter name	Probability	Kinetic parameter	Citation
Δt	N/A	$1.25\text{x}10^{-4}$ s	
Δx	N/A	30 nm	
P_{diff} (free TCR)	$8.89\text{x}10^{-4}$	0.064 $\mu m^2/s$	Chen *et al.* 2021[103]
P_{diff} (bound TCR)	$5.97\text{x}10^{-5}$	0.0043 $\mu m^2/s$	Chen *et al.* 2021[103]

2. pLAT Dephosphorylation

The activity of CD45 and other phosphatases is incorporated into the model as the dephosphorylation probability of pLAT. Since the pY sites on LAT are bound by other proteins in the cluster, the probability of dephosphorylation for an individual pLAT molecule is reduced as it forms more bonds, according to the following equation:

$$P_{dephos} = P_{max} - \frac{P_{max} - P_{min}}{1 - e^{(-B + K)}} \qquad (2)$$

Where P_{max} and P_{min} are the bounds on the probability of dephosphorylation, K is a constant that determines how many bonds are required for protection, and B describes the number and type of bonds made by the molecule:

$$B = \sum_{i=1}^{N} w_i \qquad (3)$$

Where N is the number of bonds made by the molecule and w_i is the weight of the i^{th} bond. Bond weights are high for the strong interactions mediated by PLCγ and low for the weaker interactions mediated by Grb2 and SOS. The protection from phosphatases provided by a cluster is the basis for one of several positive feedback mechanisms in condensate formation.

Table S2. pLAT dephosphorylation parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability.

Parameter name	Probability	Kinetic parameter
P_{min}	$1\text{x}10^{-6}$	$8\text{x}10^{-3}$ s^{-1}
P_{max}	$6.5\text{x}10^{-4}$	5.2 s^{-1}
K	N/A	8
w (strong bond)	N/A	8
w (weak bond)	N/A	2

3. TCR-pMHC binding and activation

TCR binding occurs through standard binding kinetics with second order kinetics for association and first order kinetics for dissociation. In each simulation, only TCRs or MHCs, but not both, are explicitly modeled, and their binding probability is modulated by a parameter which represents the 2D density of their unmodeled binding partner. In addition to being bound or free, TCRs can exist in a signaling competent or incompetent state. All TCRs start signaling competent, meaning that they can immediately activate ZAP70 and begin phosphorylating LAT as soon as they bind a pMHC. However they become signaling incompetent with first order kinetics during the times that they are bound to a pMHC. This accounts for the local negative feedback that has been observed to occur on TCRs in single molecule imaging studies[80].

Table S3. TCR-pMHC binding parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability. When a parameter was set based on a prior experimental measurement, a citation is listed for that measurement.

Parameter name	Probability	Kinetic parameter	Citation
P_{on} (always bound)	1	∞	
P_{off} (always bound)	0	0	
P_{on} (dynamic)	3.125×10^{-9}	2.5×10^{-5} μm²/(molecule x s)	McAffee *et al*. 2022[80]
P_{off} (dynamic)	2.5×10^{-6} to 2.5×10^{-3}	$0.02\ s^{-1}$ to $2\ s^{-1}$	McAffee *et al*. 2022[80]
$P_{deactivate}$	1.925×10^{-6}	$1.54 \times 10^{-2}\ s^{-1}$	

4. LAT phosphorylation

In order to focus on the contributions of LAT clustering to TCR signaling and simplify the model, active TCRs (pMHC-bound and signaling competent) are assumed to instantly phosphorylate ZAP70 which can then phosphorylate LAT. Each active TCR can phosphorylate LAT in the 3x3 grid centered on its location with a probability proportional to the number of sites in the 3x3 neighborhood that are not occupied by pLAT, representing the local substrate depletion around an active TCR. When pLAT is produced, it can be made in one of two forms representing different sites of phosphorylation. pY132 is modeled separately from the other three pY sites since it is known to behave differently. Primarily, pY132 is known to be phosphorylated more slowly by ZAP70[83,84], so pLAT is more often produced with only the other three sites phosphorylated (p_3LAT) while the fully phosphorylated form (p_4LAT) is produced more rarely. In cases where the p_4LAT phosphorylation rate is varied, the p_3LAT phosphorylation rate is varied such that the sum of the two rates is constant.

Table S4. LAT phosphorylation parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability.

Parameter name	Probability	Kinetic parameter
P_{phos} (p_3LAT)	1.5×10^{-2}	120 s^{-1}
P_{phos} (WT p_4LAT)	1.5×10^{-3}	12 s^{-1}
P_{phos} (G131D p_4LAT)	5×10^{-3}	40 s^{-1}

5. pLAT diffusion

pLAT diffusion proceeds very similarly to TCR diffusion as described in step 1, but with a faster diffusion coefficient that is modulated by the number of molecules that move together in a cluster:

$$P_{diff}(C) = \frac{D_{LAT}}{\sqrt{C}} \cdot \frac{C}{C+A} \cdot \frac{\Delta t}{(\Delta x)^2} \tag{4}$$

Where D_{LAT} is the diffusion coefficient for a single pLAT molecule, C is the number of molecules that are in the cluster, and A is the number of additional molecules that are pushed by the cluster as it moves. The first term models the reduced diffusion coefficient for larger objects embedded in the membrane. The second term accounts for the inertia of molecules not in the cluster as it collides with them. The third term accounts for the dimensions of the simulation as in equation 1.

Table S5. pLAT diffusion parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability.

Parameter name	Probability	Kinetic parameter
Δt	N/A	1.25×10^{-4} s
Δx	N/A	30 nm
D_{LAT}	N/A	0.302 $\mu m^2/s$

6. Intra-cluster rearrangement

Once in a cluster, pLAT molecules are allowed to rearrange within the cluster to model the flexibility that has been observed experimentally in LAT clusters[17,18]. This rearrangement happens at a much slower rate than free diffusion, and can only occur when the new location of the pLAT molecule allows it to make the same number of each type (PLCγ or Grb2/SOS) of bond with its new neighbors as it did before the rearrangement.

Table S6. pLAT diffusion parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability.

Parameter name	Probability	Kinetic parameter
$P_{rearrange}$	2.5×10^{-4}	1.8×10^{-3} μm^2/s

7. pLAT-pLAT binding

Adjacent molecules of pLAT can bind to each other through the implicit activity of PLCγ, Grb2, and SOS. Any molecules that are connected directly or indirectly by a bond must move together as a cluster until the bonds connecting them are broken. These bonds can either be strong bonds mediated by PLCγ, or weaker bonds mediated by Grb2 and SOS. However, since PLCγ binds preferentially to pY132 on LAT, at least one p_4LAT must participate in any strong bond. The strength of the bond is modeled by the first-order off-rate of the bond, with stronger bonds having slower off rates. In contrast, the on-rate of the bonds is modulated by their local environment to account for the local concentration of PLCγ, Grb2, and SOS. As a pLAT molecule forms more bonds, the local concentration of the crosslinking molecules is increased, thus increasing the effective on-rate of modeled pLAT-pLAT bonds. This increase on the on-rate drives a positive feedback loop which allows pLAT condensates to grow quickly once nucleated, following the equation below:

$$P_{bind}(i,j) = \left(P_{min} + \frac{P_{max} - P_{min}}{1 - e^{(-S_{ij} + K)}}\right) \cdot e^{-(E_i + E_j)} \tag{5}$$

Where P_{max} and P_{min} are the bounds on the probability of binding, K describes the degree of local binding required to promote further binding, and E_i and S_{ij} are defined as follows:
E_I is equal to 0 if molecule i is already bound to a molecule other than j, and is a positive number, Z, representing the entropic penalty of binding if molecule i is free.

$$S_{ij} = N_{ij} \cdot b \tag{6}$$

Where N_{ij} is the number of bonds made by pLAT molecules that neighbor both molecules i and j, and b is the weight of a bond. Thus, as more bonds are made around a pair of pLAT molecules, the local concentration of cross-linking molecules increases, resulting in an increased probability of binding. Because PLCγ-mediated bonds only require a single molecule to form, the P_{min} for the formation of those bonds is increased relative to the Grb2/SOS-mediated bonds, and the off rate is decreased.

Table S7. pLAT-pLAT binding parameters. Each parameter value is listed as both the probability of the event it describes happening in a given time step, and as the kinetic parameter that results from the given probability.

Parameter name	Probability	Kinetic parameter
P_{min} (weak)	$5x10^{-5}$	$0.4 s^{-1}$
P_{min} (strong)	$1x10^{-3}$	$8 s^{-1}$
P_{max} (weak or strong)	0.1	$800 s^{-1}$
K	N/A	15
b	N/A	2.5
Z	N/A	1
P_{unbind} (weak)	$4.58x10^{-4}$	$3.66 s^{-1}$
P_{unbind} (strong)	$8.3x10^{-7}$	$6.64x10^{-3} s^{-1}$

Supplemental Figures

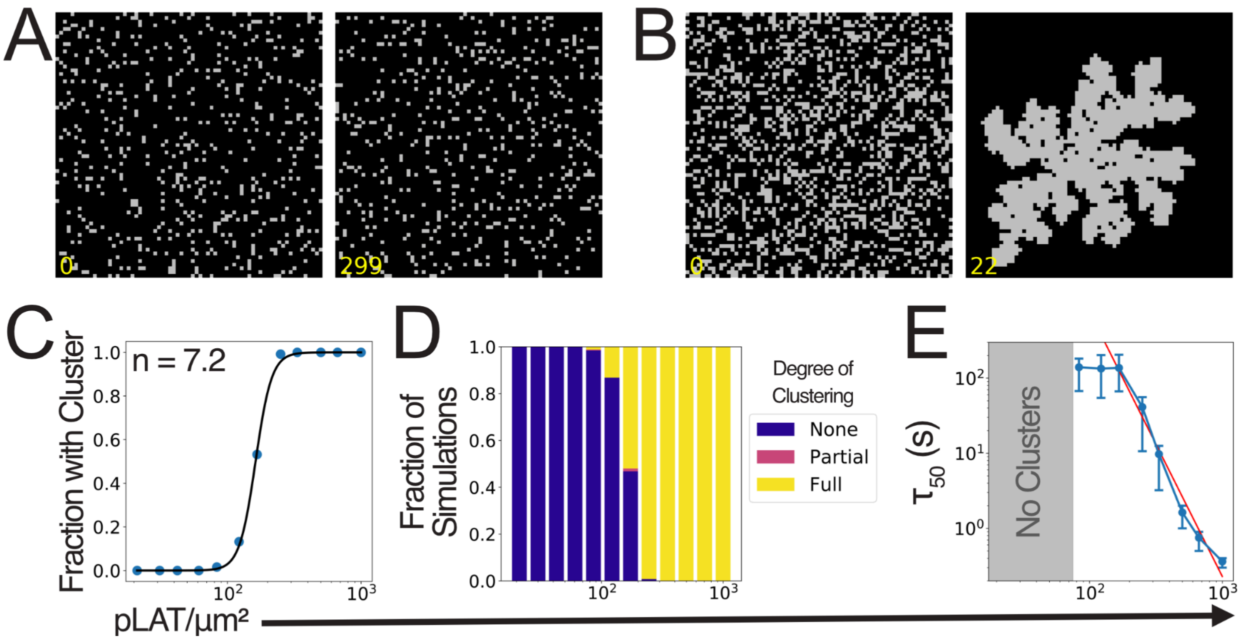

Figure S1. pLAT molecules condense rapidly above a critical concentration. Data from 250 replicate simulations without TCR or CD45 activity at varying p_3LAT 2D densities. No p_4LAT was included in these simulations. A,B) Representative snapshots of simulations with 122pLAT/µm^2 (A) or 333pLAT/µm^2 (B). Images are 2.13µm square, timestamps are min:sec. C) The fraction of simulations at each pLAT density that achieve a maximum cluster size of at least 50 molecules. Black curve represents the fit to the Hill equation. D) Stacked bar plot showing the proportion simulations that have either full, partial, or no clustering, with >95%, 5%-95%, or <5% of LAT in clusters, respectively. E) The mean and standard deviation of the time until the first cluster for each pLAT density where any clusters form. Red line shows power law fit to the points with pLAT densities achieving at least 50% clustering.

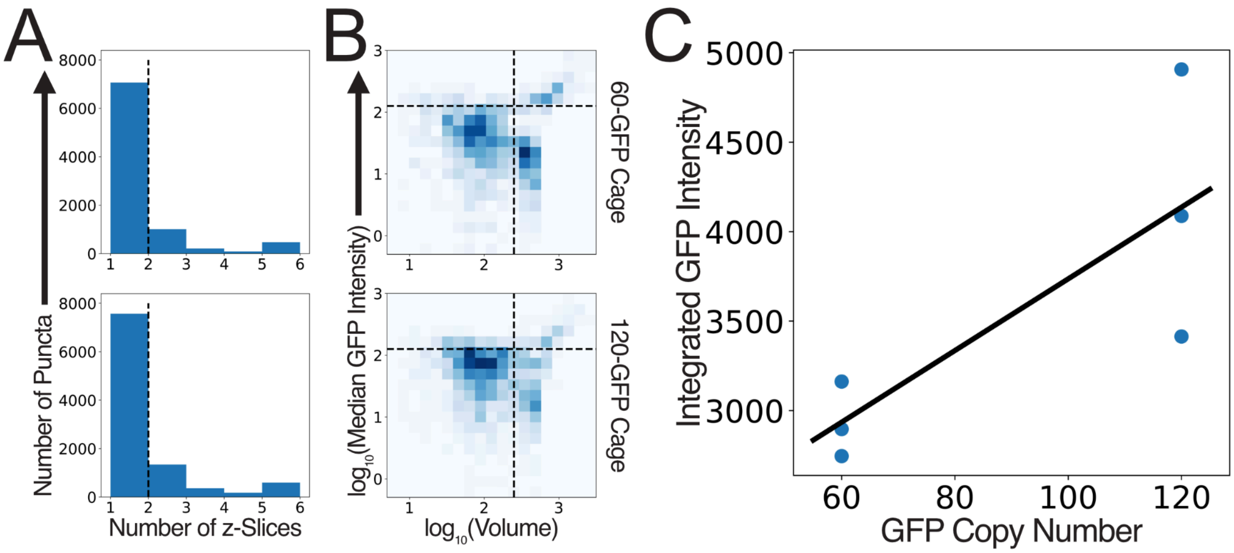

Figure S2. GFP counting calibration. A,B) Filtering criteria for inclusion of puncta in GFP counting analysis. A) Histograms showing the distribution of the number of z-slices each nanoparticle spans for the 60-GFP nanoparticle (top) and 120-GFP nanoparticle (bottom). Nanoparticles must be present in at least 2 consecutive z-slices to be included in the analysis. B) 2D histograms showing the joint distribution of the volume (voxels) and median GFP intensity (a.u.) for the 60-GFP nanoparticle (top) and 120-GFP nanoparticle (bottom). Nanoparticles must not exceed 252 voxels or a median intensity of 126 in order to be included in the analysis. C) Linear fit of the median of the integrated GFP intensity distribution of nanoparticles passing the filtering criteria to three replicates at each GFP number.

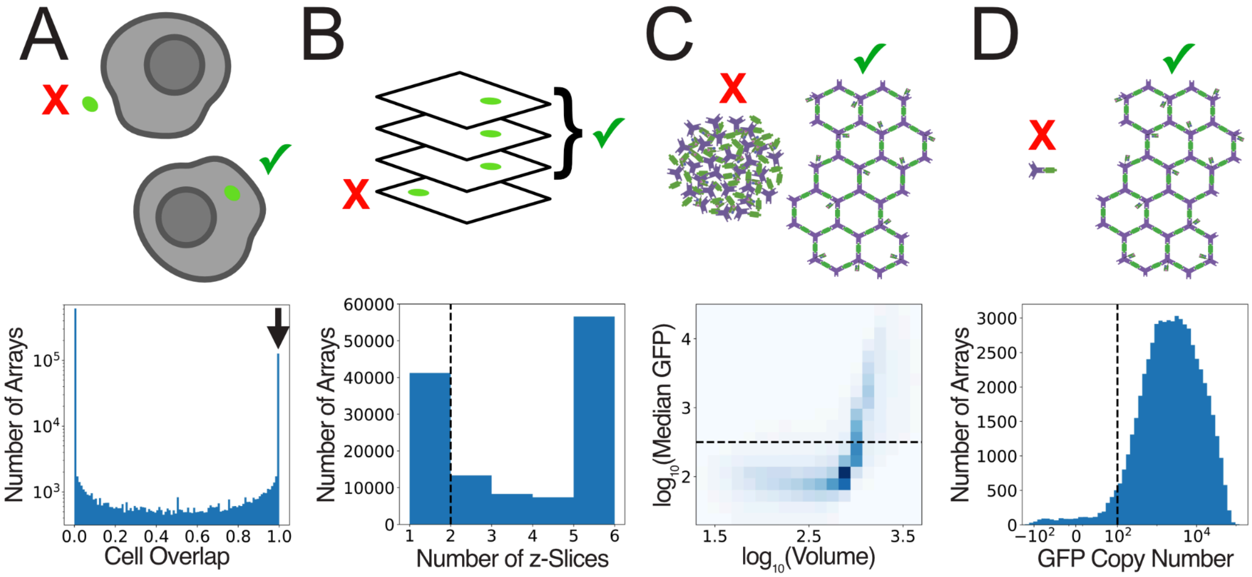

Figure S3. Filtering removes low quality arrays. Cartoons and distributions describing the filtering criteria for inclusion of arrays in later analysis. A) Cell contact filter. Histogram shows the distribution of the fraction of the array volume that overlaps with a cell for all arrays. Only arrays with at least 99% overlap are included. B) Z-continuity filter. Histogram shows the distribution of the number of consecutive z-slices the array is present in for all arrays passing the cell contact filter. Arrays must span at least 2 consecutive z-slices to be included. C) Aggregation filter. 2D histogram shows the joint distribution of the volume (voxels) and median GFP intensity (a.u.) of all arrays passing the cell contact and z-continuity filters . Arrays with median GFP greater than 317 are considered to be aggregated and are excluded. D) Size filter. Histogram shows the distribution of GFP copy number for all arrays passing the cell overlap, z-continuity, and aggregation filters. Arrays must have at least 100 GFP copies to be included.

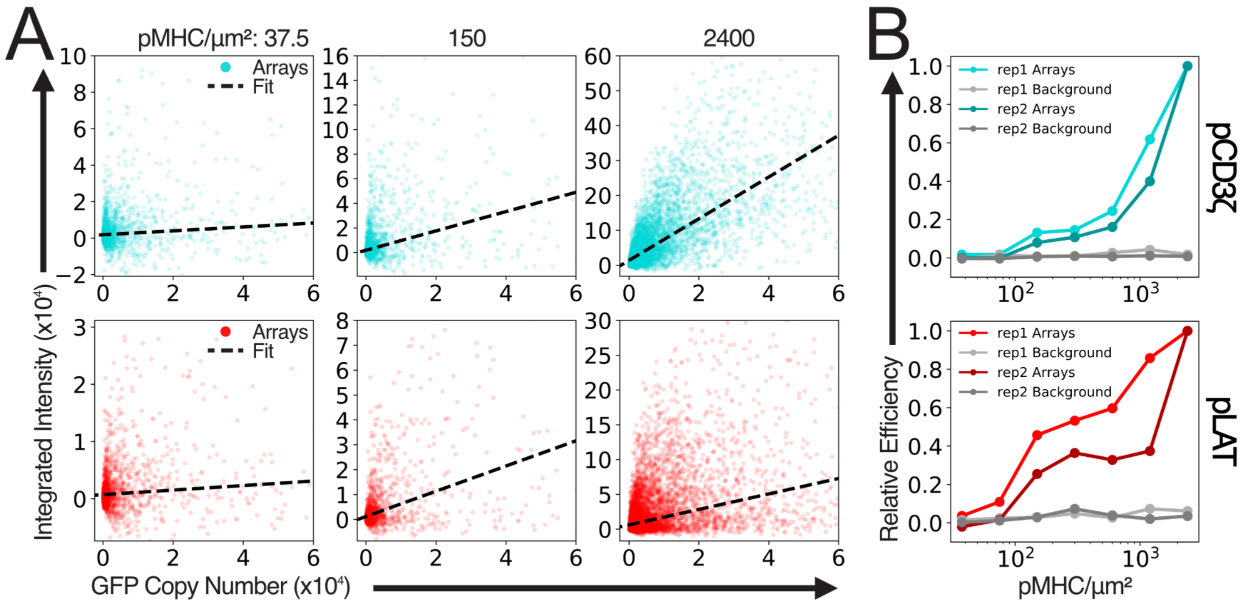

Figure S4. Total phosphorylation increases with array size and density. A) Scatterplots showing the background-subtracted integrated pCD3ζ (top) or pLAT (bottom) intensities produced by arrays vs the number of GFP molecules in the array for arrays from three representative pMHC densities. Black dashed lines indicate the line of best fit to the data. B) Line plots showing the relationship between the slope of the fitted lines in (A) and the pMHC density for two biological replicates. Each replicate is independently normalized to the value for the highest pMHC density.

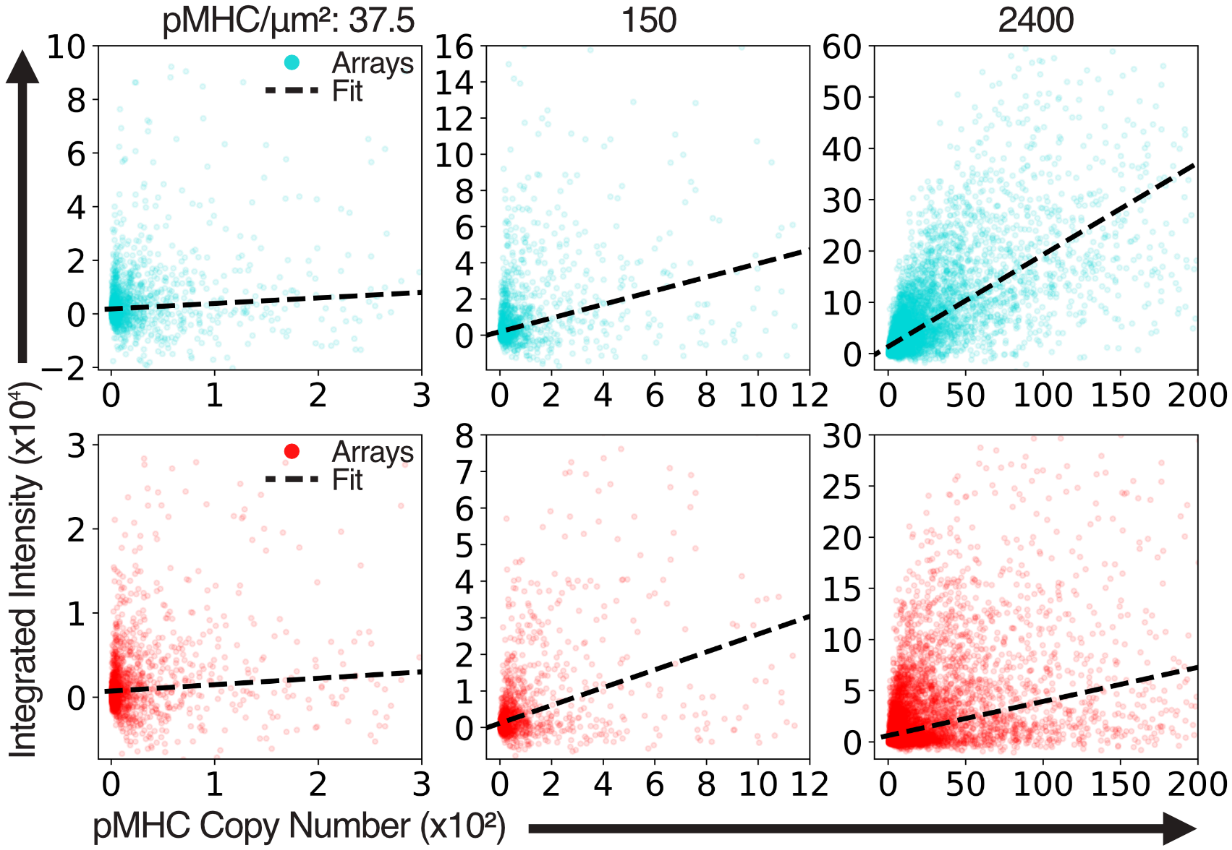

Figure S5. Phosphorylation at constant array size increases with pMHC density. Scatterplots showing the background-subtracted integrated pCD3ζ (top) or pLAT (bottom) intensities produced by arrays vs the number of pMHC molecules in the array for arrays from three representative pMHC densities. Black dashed lines indicate the line of best fit to the data.

Supplemental Movies

Supplemental movies 1-2. Simulations at constant pLAT density. Representative movies of simulations with constant p_3LAT levels showing that clustering does not occur at 122 molecules/μm^2 (movie S1), but does occur at 333 molecules/μm^2 (movie S2). Still frames from movies S1 and S2 are shown in figure S1A and S1B, respectively.

Supplemental movie 3. Simulation with constantly bound TCR. Representative movie of simulations with a TCR that is constantly bound to a pMHC showing the formation of a pLAT cluster in response to a single TCR. Still frames and a summary plot from movie S3 are shown in figure 2.

Supplemental movies 4-10. Simulations at varying pMHC off rate. Representative movies of simulations at a range of pMHC off rates, with pMHC density varied to keep TCR occupancy constant across simulations. Still frames from movies S4-10 are shown in figure 4B in order, starting with a frame from movie S4 on the right and ending with a frame from movie S10 on the left.

Supplemental movies 11-15. Simulations at varying pMHC spacing. Representative movies of simulations with a constant number of pMHCs held at varying distances from each other. Still frames from movies S11-15 are shown in figure 6B in order, starting with a frame from movie S11 on the right and ending with a frame from movie S15 on the left.

In all supplemental movies, p_3LAT is represented by light gray squares, p_4LAT is represented by dark gray squares, active TCRs are represented by red circles, and deactivated TCRs are represented by purple circles, as in the still frames in figures 2, 4, 6, and S1.

Discussion

Understanding how T-cells are able to sensitively and selectively eliminate threats from the body is critical for harnessing their behavior for a wide range of therapeutic applications across cancer, autoimmunity, and infectious disease. Here I have provided two new ways to understand and explore features of the T-cell and TCR that make these behaviors possible. SMART MHCs, if they can be expanded to cover a wider range of MHC alleles, will make the study of TCR-pMHC interactions much more tractable. By circumventing the need for refolding of soluble pMHC, SMART MHCs make T-cell staining experiments much more accessible. When refolding is not required, the equipment and expertise required to produce a pMHC reagent decreases significantly, allowing a much broader set of research groups to produce their own pMHC reagents on demand. This ease of access has the potential to dramatically accelerate any research that relies on identification, tracking, or isolation of a specific T-cell population. Additionally, SMART MHCs demonstrate the ability to improve on existing pMHC yeast display strategies. The significantly improved folding and expression levels we demonstrated likely make it possible to detect lower affinity TCR-pMHC interactions, and facilitate more careful mapping of TCR specificities. Both of these applications of SMART MHCs make them a useful tool to understand how T-cells achieve their specificity from the perspective of the structural and biochemical features that drive TCR-pMHC interactions.

Another key aspect of T-cell sensitivity and selectivity stems not from the TCR-pMHC interaction, but from the signaling cascade that this interaction initiates. Differences in TCR-pMHC affinities cannot fully explain the observed differences in signaling outcomes[22,32], but these differences can be explained by the underlying signaling network. The model presented here provides an intuitive explanation for the properties of the TCR pathway that allow it to distinguish self and foreign peptides. The nucleation and growth of LAT signaling clusters is able to explain these key properties without relying on a complex series of biochemical steps. The delay associated with nucleation allows the filtering out of interactions with self pMHCs, and the positive feedback associated with the growth of a cluster can produce single molecule sensitivity. Additionally, the length scale of cluster formation explains the preferred TCR spacing. The ability of a single biochemical step to explain this wide array of signaling properties is remarkable, and could represent a more general signaling mechanism that exists more broadly outside of TCR signaling.

Together the LAT condensation model and SMART MHCs provide important new ways to study TCR biology from two key perspectives. Understanding how both the binding specificity and the signaling specificity of TCRs arise is critical for understanding their behavior, and harnessing it for therapeutic applications. The work presented here provides a powerful tool and an important model that together lend insight into these key properties of T-cells.

Acknowledgements

Funding:
The work presented here was funded by an NIH/NIBIB Trailblazer Award (R21EB027327) given to Hao Y. Kueh, a Bill and Melinda Gates Foundation grant (INV-010680) given to David Baker, and an NIH NIAID grant (P01AI091580) given to Jay T. Groves.

Computational protein design:
I would like to thank many members of the Baker lab for their contributions to this work. In particular, Brian Coventry and Longxing Cao assisted in running the scripts necessary to design the initial library of stabilizing domains that became the SMART MHC. Kiera Sumda and Christopher Norn helped with the redesign of the cleavage sites of hit6. Joe Watson and David Jeurgens helped with the linker design that was the final step in producing SMART H-2D^b. Hua Bai contributed significantly to all steps of the computational design process.

Experimental characterization of SMART MHCs:
Would again like to thank many members of the Baker lab, and several collaborating labs for their contributions to this work. Hua Bai, Inna Goreshnik, and Aza Allen contributed to the screening and sequencing of the yeast display library of stabilizing domains. Hua Bai helped the expression and characterization of hit6 and the designs tested during the optimization process. Laura Guerrero in Michael Jewett's lab performed the CFPS expression measurements. Yi Sun and Amarachi Onwuke in Nikolaos Sgourakis' lab performed some of the peptide binding experiments, and contributed many of the MHC and peptide sequences used in this work. Renhua Sun in Adnnane Achour's lab performed the SPR measurements of pMHC-TCR binding and the crystallography for SMART H-2D^b. Tina Chen in Chris Garcia's lab performed the TCR binding yeast display experiments. Chan Kim helped with the design and execution of the SMART MHC oligomer T-cell staining experiments.

Array imaging experiments:
I would like to thank several members of the Kueh lab for their contributions to the work presented here. Matthew Wither provided the P14 Jurkat cells line, and along with Kathleen Abadie provided training in cell culture and imaging. Michael Yirdaw helped with all aspects of the imaging experiments. Additionally, Ariel Ben-Sasson in the Baker lab provided materials and protocols related to the production and assembly of the arrays.

Computing and lab resources:
Simulations were performed on the Baker lab computing cluster, and I would like to thank Luki Goldschmidt and Patrick Vecchiato for their work in maintaining the cluster. I would also like to thank the lab management of both the Baker and Kueh labs, Kandise VanWormer and Paul Leanza, for keeping the lab space clean, safe, and well-stocked. None of this work would have been possible without them.

Broad advice and input:
I would like to thank my advisors, Hoa Y. Kueh and David Baker for their continued support throughout my PhD. Both of them have contributed significant time and resources to my training and to advising this work, and it would not have been possible without them. I would also like to thank the rest of my committee for their input and feedback on my work.

References

1. Siller-Farfán, J. A. & Dushek, O. Molecular mechanisms of T cell sensitivity to antigen. *Immunol. Rev.* **285**, 194–205 (2018).
2. Kurd, N. & Robey, E. A. T-cell selection in the thymus: a spatial and temporal perspective. *Immunol. Rev.* **271**, 114–126 (2016).
3. Nagorsen, D., Scheibenbogen, C., Marincola, F. M., Letsch, A. & Keilholz, U. Natural T cell immunity against cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **9**, 4296–4303 (2003).
4. 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 (2016).
5. Birnbaum, M. E. *et al.* Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
6. Linnemann, C. *et al.* High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat. Med.* **19**, 1534–1541 (2013).
7. Dobson, C. S. *et al.* Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **19**, 449–460 (2022).
8. Sibener, L. V. *et al.* Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **174**, 672-687.e27 (2018).
9. Birnbaum, M. E., Dong, S. & Garcia, K. C. Diversity-oriented approaches for interrogating T-cell receptor repertoire, ligand recognition, and function. *Immunol. Rev.* **250**, 82–101 (2012).
10. Adams, J. J. *et al.* T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* **35**, 681–693 (2011).
11. Yokosuka, T. *et al.* Newly generated T cell receptor microclusters initiate and sustain T cell

activation by recruitment of Zap70 and SLP-76. *Nat. Immunol.* **6**, 1253–1262 (2005).

12. Pageon, S. V. *et al.* Functional role of T-cell receptor nanoclusters in signal initiation and antigen discrimination. *Proc. Natl. Acad. Sci.* **113**, E5454–E5463 (2016).
13. Li, Y.-C. *et al.* Cutting Edge: mechanical forces acting on T cells immobilized via the TCR complex can trigger TCR signaling. *J. Immunol. Baltim. Md 1950* **184**, 5959–5963 (2010).
14. Feng, Y. *et al.* Mechanosensing drives acuity of αβ T-cell recognition. *Proc. Natl. Acad. Sci.* **114**, E8204–E8213 (2017).
15. James, J. R. & Vale, R. D. Biophysical mechanism of T-cell receptor triggering in a reconstituted system. *Nature* **487**, 64–69 (2012).
16. Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).
17. Zeng, L., Palaia, I., Šarić, A. & Su, X. PLCγ1 promotes phase separation of T cell signaling components. *J. Cell Biol.* **220**, e202009154 (2021).
18. Su, X. *et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595–599 (2016).
19. Huang, W. Y. C. *et al.* A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science* **363**, 1098–1103 (2019).
20. Altan-Bonnet, G. & Germain, R. N. Modeling T Cell Antigen Discrimination Based on Feedback Control of Digital ERK Responses. *PLOS Biol.* **3**, e356 (2005).
21. Lever, M. *et al.* Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6630–E6638 (2016).
22. McKeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5042–5046 (1995).
23. François, P., Voisinne, G., Siggia, E., Altan-Bonnet, G. & Vergassola, M. Phenotypic model

for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

24. Tischer, D. K. & Weiner, O. D. Light-based tuning of ligand half-life supports kinetic proofreading model of T cell signaling. *eLife* **8**, e42498 (2019).
25. Ganti, R. S. *et al.* How the T cell signaling network processes information to discriminate between self and agonist ligands. *Proc. Natl. Acad. Sci.* **117**, 26020–26030 (2020).
26. Lipniacki, T., Hat, B., Faeder, J. R. & Hlavacek, W. S. Stochastic effects and bistability in T cell receptor signaling. *J. Theor. Biol.* **254**, 110–122 (2008).
27. Das, J. *et al.* Digital Signaling and Hysteresis Characterize Ras Activation in Lymphoid Cells. *Cell* **136**, 337–351 (2009).
28. Lin, J. J. Y. *et al.* Mapping the stochastic sequence of individual ligand-receptor binding events to cellular activation: T cells act on the rare events. *Sci. Signal.* **12**, eaat8715 (2019).
29. Kirby, D. & Zilman, A. Proofreading does not result in more reliable ligand discrimination in receptor signaling due to its inherent stochasticity. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2212795120 (2023).
30. Overall, S. A. *et al.* High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat. Commun.* **11**, 1909 (2020).
31. Wooldridge, L. *et al.* Tricks with tetramers: how to get the most from multimeric peptide–MHC. *Immunology* **126**, 147–164 (2009).
32. Pettmann, J. *et al.* The discriminatory power of the T cell receptor. *eLife* **10**, e67092 (2021).
33. Murali-Krishna, K. *et al.* Counting Antigen-Specific CD8 T Cells: A Reevaluation of Bystander Activation during Viral Infection. *Immunity* **8**, 177–187 (1998).
34. Kronenberg, D. *et al.* Circulating Preproinsulin Signal Peptide–Specific CD8 T Cells Restricted by the Susceptibility Molecule HLA-A24 Are Expanded at Onset of Type 1 Diabetes and Kill β-Cells. *Diabetes* **61**, 1752–1759 (2012).
35. Rowntree, L. C. *et al.* Preferential HLA-B27 Allorecognition Displayed by Multiple

Cross-Reactive Antiviral CD8+ T Cell Receptors. *Front. Immunol.* **11**, (2020).

36. Li, S. *et al.* Characterization of neoantigen-specific T cells in cancer resistant to immune checkpoint therapies. *Proc. Natl. Acad. Sci.* **118**, e2025570118 (2021).
37. Cohen, C. J. *et al.* Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J. Clin. Invest.* **125**, 3981–3991 (2015).
38. Lee, P. P. *et al.* Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat. Med.* **5**, 677–685 (1999).
39. Production Protocols | NIH Tetramer Core Facility. https://tetramer.yerkes.emory.edu/support/protocols#4.
40. Bakker, A. H. *et al.* Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl. Acad. Sci.* **105**, 3825–3830 (2008).
41. Luimstra, J. J. *et al.* A flexible MHC class I multimer loading system for large-scale detection of antigen-specific T cells. *J. Exp. Med.* **215**, 1493–1504 (2018).
42. Sun, Y. *et al.* Universal open MHC-I molecules for rapid peptide loading and enhanced complex stability across HLA allotypes. *Proc. Natl. Acad. Sci.* **120**, e2304055120 (2023).
43. Crawford, F., Huseby, E., White, J., Marrack, P. & Kappler, J. W. Mimotopes for Alloreactive and Conventional T Cells in a Peptide–MHC Display Library. *PLOS Biol.* **2**, e90 (2004).
44. Yang, X. *et al.* Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature* **612**, 771–777 (2022).
45. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
46. Madden, D. R. The Three-Dimensional Structure of Peptide-MHC Complexes. *Annu. Rev. Immunol.* **13**, 587–622 (1995).
47. Zareie, P. *et al.* Canonical T cell receptor docking on peptide-MHC is essential for T cell signaling. *Science* **372**, eabe9124 (2021).

48. Jones, L. L. *et al.* Engineering and Characterization of a Stabilized α1/α2 Module of the Class I Major Histocompatibility Complex Product Ld*. *J. Biol. Chem.* **281**, 25734–25744 (2006).

49. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

50. Boulter, J. M. *et al.* Potent T cell agonism mediated by a very rapid TCR/pMHC interaction. *Eur. J. Immunol.* **37**, 798–806 (2007).

51. Achour, A. *et al.* Structural Basis of the Differential Stability and Receptor Specificity of H-2Db in Complex with Murine *versus* Human β2-Microglobulin. *J. Mol. Biol.* **356**, 382–396 (2006).

52. Li, Z. *et al.* The Mechanism of β2m Molecule-Induced Changes in the Peptide Presentation Profile in a Bony Fish. *iScience* **23**, 101119 (2020).

53. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).

54. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).

55. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

56. Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).

57. Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).

58. Duru, A. D. *et al.* Tuning antiviral CD8 T-cell response via proline-altered peptide ligand vaccination. *PLOS Pathog.* **16**, e1008244 (2020).

59. Velloso, L. M., Michaëlsson, J., Ljunggren, H.-G., Schneider, G. & Achour, A. Determination of Structural Principles Underlying Three Different Modes of Lymphocytic Choriomeningitis

Virus Escape from CTL Recognition1. *J. Immunol.* **172**, 5504–5511 (2004).

60. Utz, U., Banks, D., Jacobson, S. & Biddison, W. E. Analysis of the T-cell receptor repertoire of human T-cell leukemia virus type 1 (HTLV-1) Tax-specific CD8+ cytotoxic T lymphocytes from patients with HTLV-1-associated disease: evidence for oligoclonal expansion. *J. Virol.* **70**, 843–851 (1996).
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. Chen, Y. T. *et al.* A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1914–1918 (1997).
63. McShan, A. C. *et al.* TAPBPR promotes antigen loading on MHC-I molecules using a peptide trap. *Nat. Commun.* **12**, 3174 (2021).
64. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
65. Hulme, E. C. & Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **161**, 1219–1237 (2010).
66. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
67. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
68. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
69. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
70. Berkholz, D. S., Shapovalov, M. V., Dunbrack, R. L. & Karplus, P. A. Conformation Dependence of Backbone Geometry in Proteins. *Struct. Lond. Engl. 1993* **17**, 1316–1325 (2009).

71. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci. Publ. Protein Soc.* **27**, 293–315 (2018).

72. van der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* **11**, 47–55 (2011).

73. Richard, A. C. *et al.* T cell cytolytic capacity is independent of initial stimulation strength. *Nat. Immunol.* **19**, 849–858 (2018).

74. Wülfing, C. *et al.* Kinetics and Extent of T Cell Activation as Measured with the Calcium Signal. *J. Exp. Med.* **185**, 1815–1825 (1997).

75. Wither, M. J. *et al.* Antigen perception in T cells by long-term Erk and NFAT signaling dynamics. *Proc. Natl. Acad. Sci.* **120**, e2308366120 (2023).

76. Brownlie, R. J. & Zamoyska, R. T cell receptor signalling networks: branched, diversified and bounded. *Nat. Rev. Immunol.* **13**, 257–269 (2013).

77. Courtney, A. H., Lo, W.-L. & Weiss, A. TCR Signaling: Mechanisms of Initiation and Propagation. *Trends Biochem. Sci.* **43**, 108–123 (2018).

78. Di Bartolo, V. *et al.* Tyrosine 319, a Newly Identified Phosphorylation Site of ZAP-70, Plays a Critical Role in T Cell Antigen Receptor Signaling*. *J. Biol. Chem.* **274**, 6285–6294 (1999).

79. Yi, J., Balagopalan, L., Nguyen, T., McIntire, K. M. & Samelson, L. E. TCR microclusters form spatially segregated domains and sequentially assemble in calcium-dependent kinetic steps. *Nat. Commun.* **10**, 277 (2019).

80. McAffee, D. B. *et al.* Discrete LAT condensates encode antigen information from single pMHC:TCR binding events. *Nat. Commun.* **13**, 7446 (2022).

81. Sun, S., GrandPre, T., Limmer, D. T. & Groves, J. T. Kinetic frustration by limited bond availability controls the LAT protein condensation phase transition on membranes. *Sci. Adv.* **8**, eabo5295 (2022).

82. Ben-Sasson, A. J. *et al.* Design of biologically active binary protein 2D materials. *Nature* **589**, 468–473 (2021).

83. Lo, W.-L. *et al.* Slow phosphorylation of a tyrosine residue in LAT optimizes T cell ligand discrimination. *Nat. Immunol.* **20**, 1481–1493 (2019).
84. Lo, W.-L. *et al.* A single-amino acid substitution in the adaptor LAT accelerates TCR proofreading kinetics and alters T-cell selection, maintenance and function. *Nat. Immunol.* **24**, 676–689 (2023).
85. Britain, D. M., Town, J. P. & Weiner, O. D. Progressive enhancement of kinetic proofreading in T cell antigen discrimination from receptor activation to DAG generation. *eLife* **11**, e75263 (2022).
86. Yousefi, O. S. *et al.* Optogenetic control shows that kinetic proofreading regulates the activity of the T cell receptor. *eLife* **8**, e42475 (2019).
87. Zhang, W. *et al.* Association of Grb2, Gads, and Phospholipase C-γ1 with Phosphorylated LAT Tyrosine Residues: EFFECT OF LAT TYROSINE MUTATIONS ON T CELL ANTIGEN RECEPTOR-MEDIATED SIGNALING*. *J. Biol. Chem.* **275**, 23355–23361 (2000).
88. Shah, N. H. *et al.* An electrostatic selection mechanism controls sequential kinase signaling downstream of the T cell receptor. *eLife* **5**, e20105 (2016).
89. Brownlie, R. J. & Zamoyska, R. T cell receptor signalling networks: branched, diversified and bounded. *Nat. Rev. Immunol.* **13**, 257–269 (2013).
90. Cho, S. *et al.* Structural basis for differential recognition of tyrosine-phosphorylated sites in the linker for activation of T cells (LAT) by the adaptor Gads. *EMBO J.* **23**, 1441–1451 (2004).
91. Courtney, A. H. *et al.* CD45 functions as a signaling gatekeeper in T cells. *Sci. Signal.* **12**, eaaw8151 (2019).
92. Case, L. B., Ditlev, J. A. & Rosen, M. K. Regulation of Transmembrane Signaling by Phase Separation. *Annu. Rev. Biophys.* **48**, 465–494 (2019).
93. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).

94. Sommers, C. L. *et al.* A LAT Mutation That Inhibits T Cell Development Yet Induces Lymphoproliferation. *Science* **296**, 2040–2043 (2002).

95. Gottschalk, R. A. *et al.* Distinct influences of peptide-MHC quality and quantity on in vivo T-cell responses. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 881–886 (2012).

96. Cai, H. *et al.* Full control of ligand positioning reveals spatial thresholds for T cell receptor triggering. *Nat. Nanotechnol.* **13**, 610–617 (2018).

97. Hellmeier, J. *et al.* DNA origami demonstrate the unique stimulatory power of single pMHCs as T cell antigens. *Proc. Natl. Acad. Sci.* **118**, e2016857118 (2021).

98. Dong, R. *et al.* DNA origami patterning of synthetic T cell receptors reveals spatial control of the sensitivity and kinetics of signal activation. *Proc. Natl. Acad. Sci.* **118**, e2109057118 (2021).

99. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109**, E690–E697 (2012).

100. Pircher, H. *et al.* Molecular analysis of the antigen receptor of virus-specific cytotoxic T cells and identification of a new Vα family. *Eur. J. Immunol.* **17**, 1843–1846 (1987).

101. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).

102. Wilhelm, K. B. *et al.* Height, but not binding epitope, affects the potency of synthetic TCR agonists. *Biophys. J.* **120**, 3869–3880 (2021).

103. Chen, K. Y. *et al.* Trapping or slowing the diffusion of T cell receptors at close contacts initiates T cell signaling. *Proc. Natl. Acad. Sci.* **118**, e2024250118 (2021).