PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Design of solubly expressed miniaturized SMART MHCs

William L. White[a,b,1,2], Hua Bai[b,c,d,3], Chan Jhong Kim[b,c], Kevin M. Jude[e], Renhua Sun[f], Laura Guerrero[g], Xiao Han[f], Xiaojing Chen[e], Apala Chaudhuri[h,4], Julia E. Bonzanini[a,b,c], Yi Sun[h,i], Amarachi E. Onwuka[h,i], Nan Wang[e,j], Chunyu Wang[e], Per-Åke Nygren[k,l], Xinting Li[b,c], Inna Goreshnik[b,c], Aza Allen[b,c], Paul M. Levine[b,c], Hao Yuan Kueh[a,k], Michael C. Jewett[g], Nikolaos G. Sgourakis[h,i], Adnane Achour[f], K. Christopher Garcia[e,j,m], and David Baker[b,c,d,1]

Affiliations are included on p. 9.

Edited by Stephen Mayo, California Institute of Technology, Pasadena, CA; received March 15, 2025; accepted December 4, 2025

The precise recognition of specific peptide–major histocompatibility complex (pMHC) complexes by T cell receptors (TCRs) plays a key role in infectious disease, cancer, and autoimmunity. A critical step in many immunobiological studies is the identification of T cells expressing TCRs specific to a given pMHC antigen. However, the intrinsic instability of empty class-I MHCs limits their soluble expression in *Escherichia coli* and makes it very difficult to characterize even a small fraction of possible pMHC/TCR interactions. To overcome this limitation, we designed small proteins which buttress the peptide binding groove of class I MHCs, replacing β2-microglobulin (β2m) and the heavy chain α3 domain, and enable soluble and partially soluble expression in *E. coli* of H-2D^b and A*02:01, respectively. We demonstrate that these soluble, monomeric, antigen-receptive, truncated (SMART) MHCs retain both peptide- and TCR-binding specificity and that peptide-bound structures of both allomorphs are similar to their full-length, native counterparts. With extension to the majority of HLA alleles, SMART MHCs should be broadly useful for probing the T cell repertoire in approaches ranging from yeast display to T cell staining.

protein design | major histocompatibility complex | soluble expression | T cell receptor

Recombinantly expressed peptide–major histocompatibility complexes (pMHCs) are widely used as staining reagents to identify or isolate T cell subsets that recognize a peptide of interest (1). They are often used to study T cell specificity (2–4), infectious disease (5), autoimmunity (6, 7), and cancer immunology (8–10). Recombinant pMHCs are also critical in determining the structures of peptide/MHC and pMHC/T cell receptor (TCR) complexes (2, 11, 12). All these discoveries were made despite the significant difficulties involved in producing the soluble pMHCs necessary for the underlying biophysical, structural, and functional experiments. The pMHC production process, which involves separate expression in *Escherichia coli* of the two MHC chains as insoluble inclusion bodies, solubilization, and refolding in the presence of the desired peptide (13), is expensive, slow, and inefficient. To reduce the burden of refolding, systems have been developed wherein a single refolding reaction is split and loaded with many different peptides (14–17); these methods have increased the number of peptide variants that can be studied, but are still limited by the need for refolding. Eukaryotic expression systems that fold the MHC structure natively (18–22) have been used in peptide library screens where peptide variants are fused to the MHC and displayed on the cell surface, but can be limited in library size or expression levels.

The difficulties in producing pMHCs likely stem from the inherent instability of the MHC molecule when either a peptide or the β2m subunit is absent. A more stable MHC-like molecule that could be readily expressed in *E. coli* or yeast would enable the study of peptide-specific T cell populations at a much larger scale. Instead of relying on specialized facilities to produce refolded pMHCs for staining experiments (13), immunologists could produce them in-house, dramatically improving their ability to iterate through multiple peptide variants or MHC alleles. A stabilized native-like MHC could also 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 (23) to replace portions of the MHC with a designed protein scaffold. We set out to design soluble, monomeric, antigen-receptive, truncated (SMART) MHC molecules that replace the α3 domain and the β2m subunit with a small designed protein domain that buttresses the peptide binding groove of the pMHC. These stabilizing domains should preserve the native peptide- and TCR-binding properties, and allow soluble expression in the absence of a peptide

Significance

Despite the critical role T cell receptor (TCR)/peptide–major histocompatibility complex (pMHC) interactions play in human health, it remains difficult to produce reagents necessary to study them. Requirements for refolding or sequence optimization limit immunologists' and biochemists' ability to characterize diverse pMHC/TCR interactions. We developed a de novo designed protein domain that stabilizes H-2D^b and partially stabilizes A*02:01 class I MHC allomorphs, allowing soluble expression in *Escherichia coli* without the need for a stabilizing peptide and improving display on the yeast surface, while maintaining peptide and TCR binding interactions. These features facilitate a wide range of experiments to more fully understand the nature of pMHC/TCR interactions and lay the foundation for the development of stabilizing domains for all MHC allomorphs.

Preprint servers: An earlier draft of this manuscript was uploaded to bioRxiv: https://www.biorxiv.org/content/10.1101/2025.03.14.643101v1.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: wlwhite.mit@gmail.com or dabaker@uw.edu.

[2]Present address: Department of Computer Science, Tufts University, Medford, MA 02155.

[3]Present address: Director of Protein Design, Lila Biologics, Inc., Seattle, WA 98102.

[4]Present address: Metabolite Scientist, Olaris, Inc., Framingham, MA 01702.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2505932123/-/DCSupplemental.

Published January 2, 2026.

without the need for refolding, providing a peptide-receptive MHC that can be loaded with arbitrary peptides.

Design of SMART MHCs

Native class I MHCs are composed of a heavy chain, the β2m subunit, and a peptide (24). The heavy chain α1 and α2 domains consist of a β-sheet supporting two α-helices which create a peptide binding groove that defines the peptide binding specificity of each MHC-I allomorph and facilitates interactions with TCRs (24). The heavy chain α3 domain and the β2m subunit are more membrane proximal and function as structural support for the α1 and α2 domains (24). The α3 domain additionally provides a binding site for the CD8 coreceptor on T cells (25). Since class I pMHC structures are well conserved, we selected the mouse H-2D^b allele as a representative in our stabilizing design process (Fig. 1*A*).

We first removed as much of the structure as possible without disrupting the ability of the MHC to present peptides and interact with TCRs. Following the precedent of previously published truncated "mini-MHC" systems, we removed the α3 and β2m domains completely (26, 27) (Fig. 1 *B*, *Left*). The truncated MHC heavy chain (residues 1 to 179) contains a hydrophobic patch on the underside of its β-sheet which we targeted for our stabilizing domain design. We used methods developed for protein binder design (23) to create this de novo stabilizing domain. We collected a set of candidate backbone structures for the stabilizing domain (Fig. 1 *B*, *Left*), docked these backbones against the truncated H-2D^b structure, and designed their sequences to create favorable contacts with the MHC, replacing the interactions made by the deleted α3 domain and β2m subunit (Fig. 1 *B*, *Middle*). Finally, we linked each designed stabilizing domain to the N terminus of the truncated MHC by a poly-GGS linker (Fig. 1 *B*, *Right*).

We screened a total of about 10^4 designs generated by this method using yeast surface display (28). We sorted for designs that enabled the expression of truncated H-2D^b on the yeast surface (*SI Appendix*, Fig. S1*A*) and were able to bind to a FITC-labeled gp33 peptide (FITC-gp33) (*SI Appendix*, Fig. S1*B*), which is bound strongly by native H-2D^b (29, 30). Expression sorting provided only slight separation of our designs from sequence-scrambled negative controls, while peptide sorting created clear separation between successful designs and controls (Fig. 1 *C* and *D*). Interestingly, designs containing a Trp residue placed similarly to W60 of β2m in the native H-2D^b/β2m/gp33 structure were positively enriched, aligning with previous observations that W60 plays a critical role in the interaction between β2m and the MHC heavy chain (31, 32).

We identified the 30 designs with the highest peptide-binding enrichment and expressed them in *E. coli*. The best performing

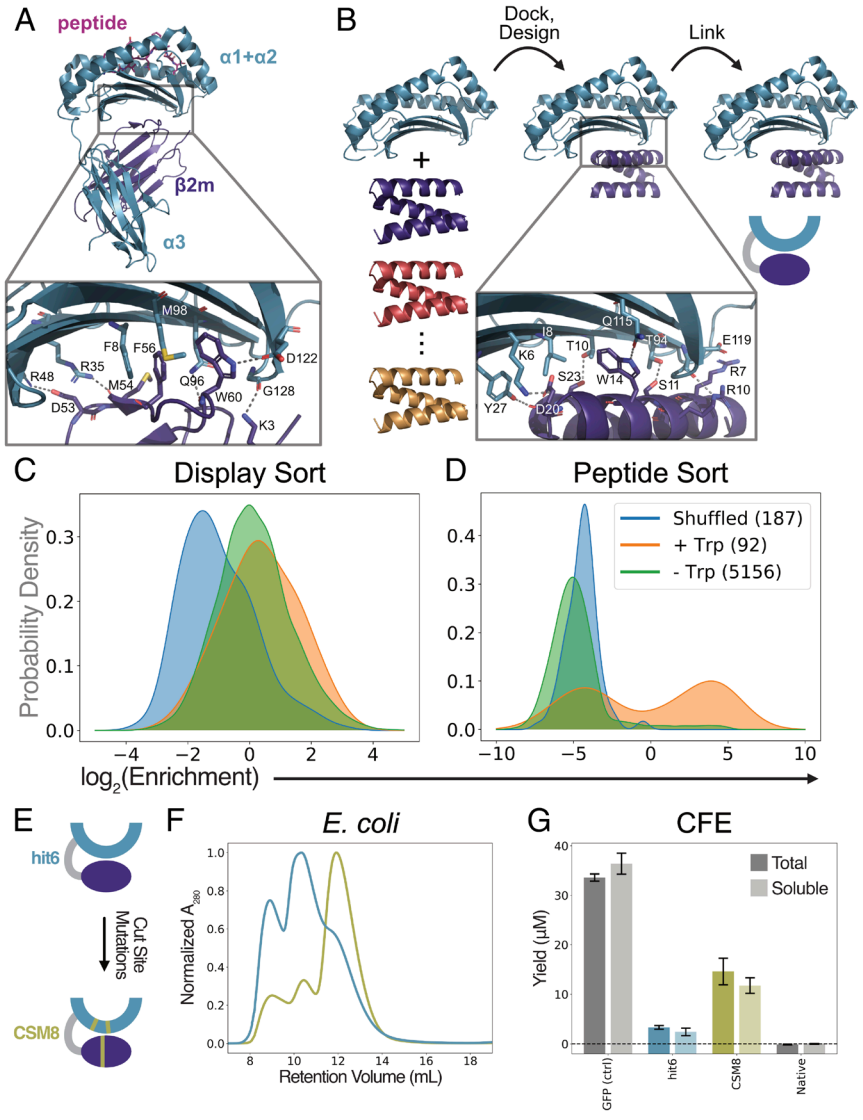

Fig. 1. Design and screening of SMART MHCs. (*A*) Structure of native H-2D^b/β2m/gp33 complex. The *Inset* shows selected interactions that are formed between the β2m subunit and the α1 and α2 domains. (*B*) Schematic of the initial design process showing the truncated H-2D^b with candidate stabilizers (*Left*), the placement of the stabilizing domain (*Middle*), and the final linked version used for yeast display (*Right*). The inset shows selected interactions between the stabilizing domain and the α1 and α2 domains. (*C* and *D*) Kernel density estimates (KDEs) of the distribution of enrichment values of all designs in either the expression (*C*) or peptide binding (*D*) yeast display sorts. Designs are split into three categories for KDE calculations: sequence-shuffled controls (shuffled), designs that do not replace W60 from β2m (−Trp), and designs that do replace W60 (+Trp). Relative enrichment was calculated by dividing design abundances after the indicated sort by abundances prior to that sort. (*E*) Schematic of the iterative improvements from hit6 to the final SMART design, showing the cleavage site mutations and the redesigned linker. (*F*) SEC traces of soluble material from *E. coli* expression cultures. (*G*) Total and soluble yields from CFE reactions for GFP (positive control) and different H-2D^b variants.

design (hit6; Fig. 1 *E*, *Top*) enabled partial folding of empty MHCs in both yeast and *E. coli* expression systems. However, in *E. coli,* only a fraction of the hit6 protein was soluble and was susceptible to proteolysis, resulting in very low yields (Fig. 1*F* and *SI Appendix*, Table S7). To improve hit6, we identified likely cleavage sites based on the masses of the proteolytic fragments, and redesigned the amino acid sequence near those sites, keeping the amino acids that directly interact with the peptide or TCR fixed, and biasing toward amino acids that occur frequently in other MHC alleles (*SI Appendix*, Fig. S2). We chose the cleavage site mutant (CSM) with the lowest degree of proteolysis: CSM8 (Fig. 1 *E*, *Bottom*). To distinguish between these variants, data in all figures are colored teal or yellow to indicate hit6 or CSM8, respectively.

We tested the expression of the SMART constructs in a cell-free expression (CFE) system compatible with a variety of high-throughput screening methods (33–41). Engineered CFE systems can produce high protein yields (39) and create oxidizing environments for disulfide bond formation, which is needed for MHC folding (40–42). The CFE results confirmed the *E. coli* results, demonstrating that, in contrast to the native H-2D^b, the stabilized designs can be expressed solubly, and that CSM8 H-2D^b is expressed more efficiently than hit6 H-2D^b (Fig. 1*G*). These results indicate that SMART MHCs can be expressed in a variety of systems that are not compatible with soluble expression of native MHC-I molecules.

SMART H-2D^b Retains Native Binding Properties and Structure

To verify that the soluble, peptide-free CSM8 H-2D^b material produced in *E. coli* retained the functional characteristics of the native MHC, we first measured the binding affinity of CSM8 H-2D^b for FITC-gp33 using fluorescence polarization (FP). These measurements demonstrated that CSM8 H-2D^b binds FITC-gp33 with an apparent dissociation constant ($K_{D,\ app}$) below 1 nM (Fig. 2*A*); this value is lower than the previously reported value of 21 nM for native H-2D^b (43, 44), likely because a competitor peptide was included in the native measurement.

Next, we assessed the binding affinity of the H-2D^b/gp33-specific P14 TCR to CSM8 H-2D^b loaded with three variants of the gp33 peptide that bind to H-2D^b with similar affinity but altered recognition by the P14 TCR (29, 45) (Fig. 2*B*). Surface plasmon resonance (SPR) measurements revealed TCR binding affinities similar to native H-2D^b/peptide complexes for all three variants (45) (Table 1, Fig. 2*C*, and *SI Appendix*, Fig. S3), further indicating that our stabilizing domain maintains the peptide-binding domain and peptide in a native-like conformation.

As a final validation of our design, the crystal structure of the CSM8 H-2D^b/gp33 complex was determined to 2.0 Å resolution. This high resolution allowed us to analyze the interactions formed between the peptide binding cleft and the stabilizing domain underneath, as well as to compare the conformation of the presented peptide, and of residues known to be essential for TCR recognition, with the native H-2Db molecule presenting the same epitope (30, 46). The structures of the peptide binding groove and full CSM8 H-2D^b design closely matched the native structure and design model, with Cα-RMSDs of 0.60 Å over 176 atoms, and 0.70 Å over 233 atoms, respectively (Fig. 2 *D* and *E*). Examination of the region surrounding the β2m tryptophan residue W60 revealed that the Trp side chain takes a similar conformation in the CSM8 design model and both crystal structures (Fig. 2*F*). Furthermore, the peptide backbone aligns well with the native conformation, and the side chains display only minor

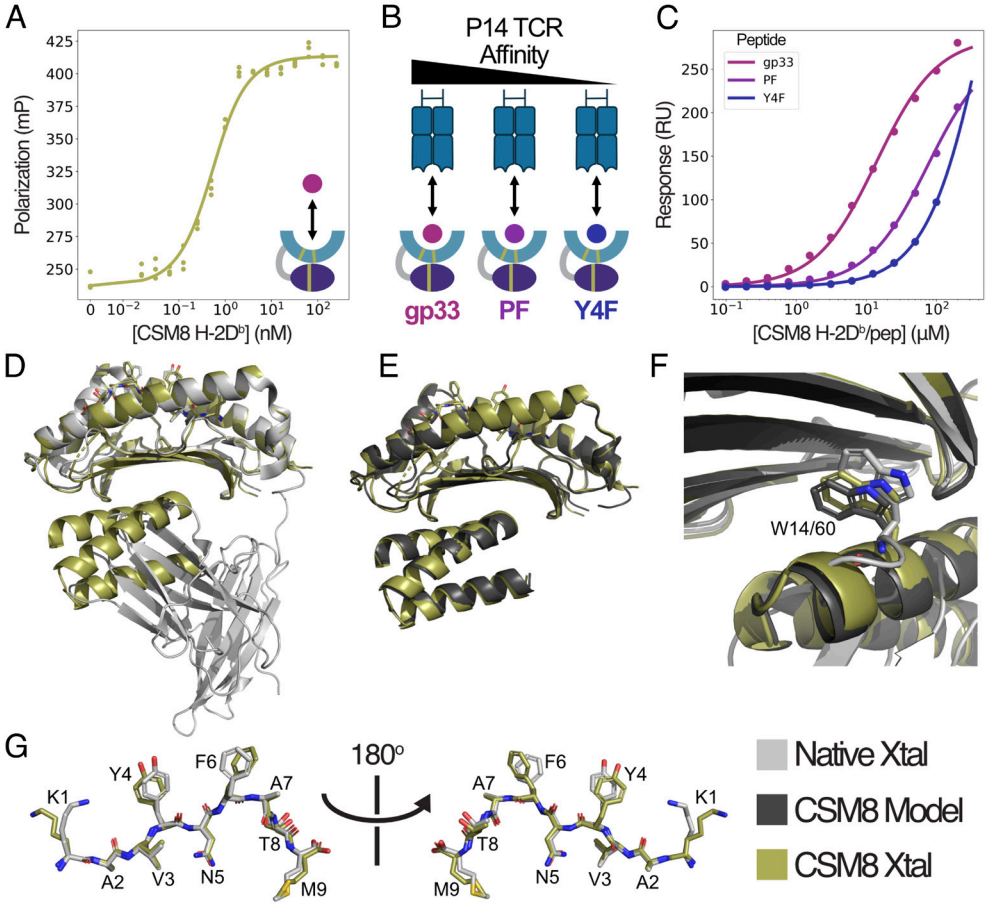

Fig. 2. SMART H-2D^b closely recapitulates the structure and interactions of the native MHC peptide binding domain. (*A*) FP data (filled circles) and fitted binding curves (lines) for FITC-gp33 binding to purified SMART H-2D^b. Each point represents one of n = 3 technical replicates. (*B*) Cartoon representations of gp33 peptide variants with varying affinity for the P14 TCR. (*C*) SPR 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 superposition of the crystal structure of the CSM8 H-2D^b/gp33 complex (yellow; PDB ID: 9HY4) onto the previously determined H-2D^b/β2m/gp33 complex (light gray; PDB ID: 1S7U). (*E*) Structural alignment of the CSM8 H-2D^b/gp33 complex crystal structure (yellow) to the design model (dark gray). (*F*) Structural alignment of all structures from (*D* and *E*) highlighting the similar positioning of W60 in the native structure relative to W14 of the CSM8 stabilizing domain. (*G*) Comparison of the conformation of the gp33 peptide in the two crystal structures.

Table 1. P14 TCR binding affinities of CSM8 H-2D^b in complex with three gp33 peptide variants

Peptide name	Peptide sequence	CSM8 H-2D^b affinity*	Native H-2D^b affinity†
gp33	KAVYNFATM	14.1 ± 0.8 μM	8.6 ± 0.4 μM
PF	KA**PF**NFATM	75.6 ± 4.3 μM	80 ± 22 μM
Y4F	KAV**F**NFATM	609 ± 64 μM	No binding detected

CSM8 affinities are calculated by fitting a standard binding curve to the equilibrium binding SPR values. Mutations relative to the gp33 sequence are shown in bold.
*Values are reported as fitted KD ± 1 SD (as determined curve-fitting covariance).
†Native affinities were measured previously by Duru et al. (45).

changes between the two crystal structures (Fig. 2*G*). These small changes, along with a few minor shifts elsewhere in the CSM8 H-2D^b structure could explain the slight deviations in binding affinities observed in our SPR experiments. Overall, these results demonstrate that CSM8 H-2D^b retains the structural features that allow peptide and TCR binding, without the need for refolding.

CSM8 HLA A*02:01 Displays Partial Solubility and Stability

Next, we tested the generalizability of our stabilizing domain to other MHC allomorphs. Based on the structural and sequence similarity of H-2D^b to many HLAs we did not introduce any modifications to the stabilizing domain, varying only the MHC sequence. We selected HLA A*02:01 due to its high frequency across different ethnic groups (47), and the large number of well-characterized peptide epitopes and TCRs that bind to it (48, 49). Although the CSM8 A*02:01 construct could be expressed in a soluble form in *E. coli*, a portion of the soluble material was in a dimeric (and likely misfolded) state, which reappeared after isolation of the monomeric fraction (Fig. 3*A*).

To test the ability of CSM8 A*02:01 to present peptides, we selected the tumor-associated antigen, NY-ESO-1 (50) and the well-characterized NY-ESO-1/A*02:01-specific TCR, 1G4 (49). We used FP methods very similar to those used to evaluate SMART H-2D^b to measure the $K_{D,\,app}$ of NY-ESO-1 to CSM8 A*02:01 and found it to be well above 1 μM (Fig. 3*B*), much weaker than the mid-nM affinities measured for the native A*02:01 using competition binding assays (51, 52). The TCR binding affinity was similarly impacted, resulting in roughly 15-to-20-fold weaker binding than reported values for native A*02:01 (49) (Table 2 and *SI Appendix*, Fig. S4 *A* and *B*).

We reasoned that these differences in affinity resulted from a lowered effective concentration due misfolding and dimerization. Therefore, we prepared CSM8 A*02:01/NY-ESO-1 complexes by standard refolding procedures and measured their affinity for the 1G4 TCR. We found that these affinities were much closer to the expected native affinities (Table 2, Fig. 3*C*, and *SI Appendix*, Fig. S4 *C* and *D*), although they were still weaker by roughly twofold to threefold. The improvement in affinity achieved by refolding, and the similar ratio of affinities between the two peptide variants across all formats, indicates that CSM8 A*02:01 can maintain a native-like structure, but only for a small fraction of molecules in solution. Overall, these data suggest that our stabilizing domain generalizes weakly to new allotypes, providing limited stabilization and solubility to A*02:01.

The SMART Stabilizer Facilitates Yeast Display of A*02:01

As noted above, yeast display can be a powerful tool to screen libraries of peptide variants for TCR binding (19, 22). Thus, we evaluated the ability of SMART A*02:01 to express on the yeast surface and bind a known TCR. We compared hit6 and CSM8 versions to assess the impact of the cleavage site mutations in this expression system. We fused the TAX9 peptide (53) to both variants, including the W167A MHC mutation to accommodate the peptide linker (19) (*SI Appendix*, Fig. S5*B*). We compared the SMART A*02:01 designs to native A*02:01 single-chain trimers (SCT) (*SI Appendix*, Fig. S5*A*) and found that it is displayed at higher levels, and achieved similar levels of binding to the high-affinity A6c134 TCR (48, 53) without binding to unrelated TCR or streptavidin (SA) controls (Fig. 3 *D* and *E* and *SI Appendix*, Fig. S5). Hit6 but not CSM8 was stained by an anti-A*02:01 antibody, suggesting that both variants are properly folded, but the cleavage site mutations prevent antibody recognition. Thus, SMART A*02:01 enhances yeast expression levels of HLA A*02:01 while retaining TCR binding specificity.

CSM8 A*02:01 Presents TAX9 Peptide to a TCR in a Native-Like Manner

To investigate whether CSM8 A*02:01 presents peptides and interacts with TCR in a native-like way, we determined the crystal structure of CSM8 A*02:01/TAX9 complexed with the A6c134 TCR. As in our SPR experiments with this allotype, refolding methods were necessary to make sufficient quantities of monomeric CSM8 A*02:01 for crystallography. We found excellent agreement between CSM8 and native A*02:01 within the HLA (Cα-RMSD 0.40 Å over 161 atoms) and TAX9 peptide (all-atom RMSD 0.51 Å) (Fig. 4*A*). CSM8 A*02:01 conserves the extensive hydrogen bond and hydrophobic contact network to the C and N termini of the TAX9 peptide (Fig. 4*B* and *SI Appendix*, Fig. S6).

We further found that the TCR variable domains engage CSM8 A*02:01 nearly identically to native A*02:01 [Cα-RMSD 0.54 Å for 69 atoms in the complementarity determining region (CDR) loops]. Though not all side chains are clearly visible in the electron density, those that are observed suggest that most or all residue contacts from the native complex are retained. In particular, we observe hydrogen bonds from the TCR β chain to the α2 helix of CSM8 A*02:01 including Glu102$^{\beta}$ to H221 and Ala101$^{\beta}$ to Ala220. At the α2 helix, the TCR α chain contributes hydrogen bonds from Asp99$^{\alpha}$ and Thr98$^{\alpha}$ to Arg135 and Gln30$^{\alpha}$ to K136 and also an extensive hydrophobic interaction between Trp101$^{\alpha}$ and the surface formed by Gln132, Ala139, and Lys138 (Fig. 4*C*). A6c134 also makes four hydrogen bonds to TAX9: Glu30$^{\beta}$ to Tyr8, Ser100$^{\alpha}$ to Gly4, Ser31$^{\alpha}$ to Tyr5, and a main-chain–mainchain contact between Ser100$^{\alpha}$ and Gly4 (Fig. 4*D*). The peptide conformation in the CSM8 and native A*02:01/TAX9/A6c134 crystal structures is conserved, with an all-atom RMSD of 0.479 Å (Fig. 4*E*). The close agreement of the CSM8 and native structures, in conjunction with peptide and TCR binding data, suggests that, despite issues with soluble expression, CSM8 A*02:01 can present peptides in the same conformation as native A*02:01.

Rigid Linker Design Improves Expression and Maintains Peptide and TCR Binding

Although we were able to purify soluble material for SMART versions of both H-2D^b and A*02:01, yields were lower than desired (*SI Appendix*, Table S7), likely due to aggregation during expression. We hypothesized that this aggregation was due, in part, to the flexibility of the poly-GGS linker connecting the stabilizer to the MHC. Thus, we used inpainting, ProteinMPNN, and AlphaFold2 (54–56) to generate rigid linkers to replace the flexible linker in CSM8. We refer to the most soluble of these improved

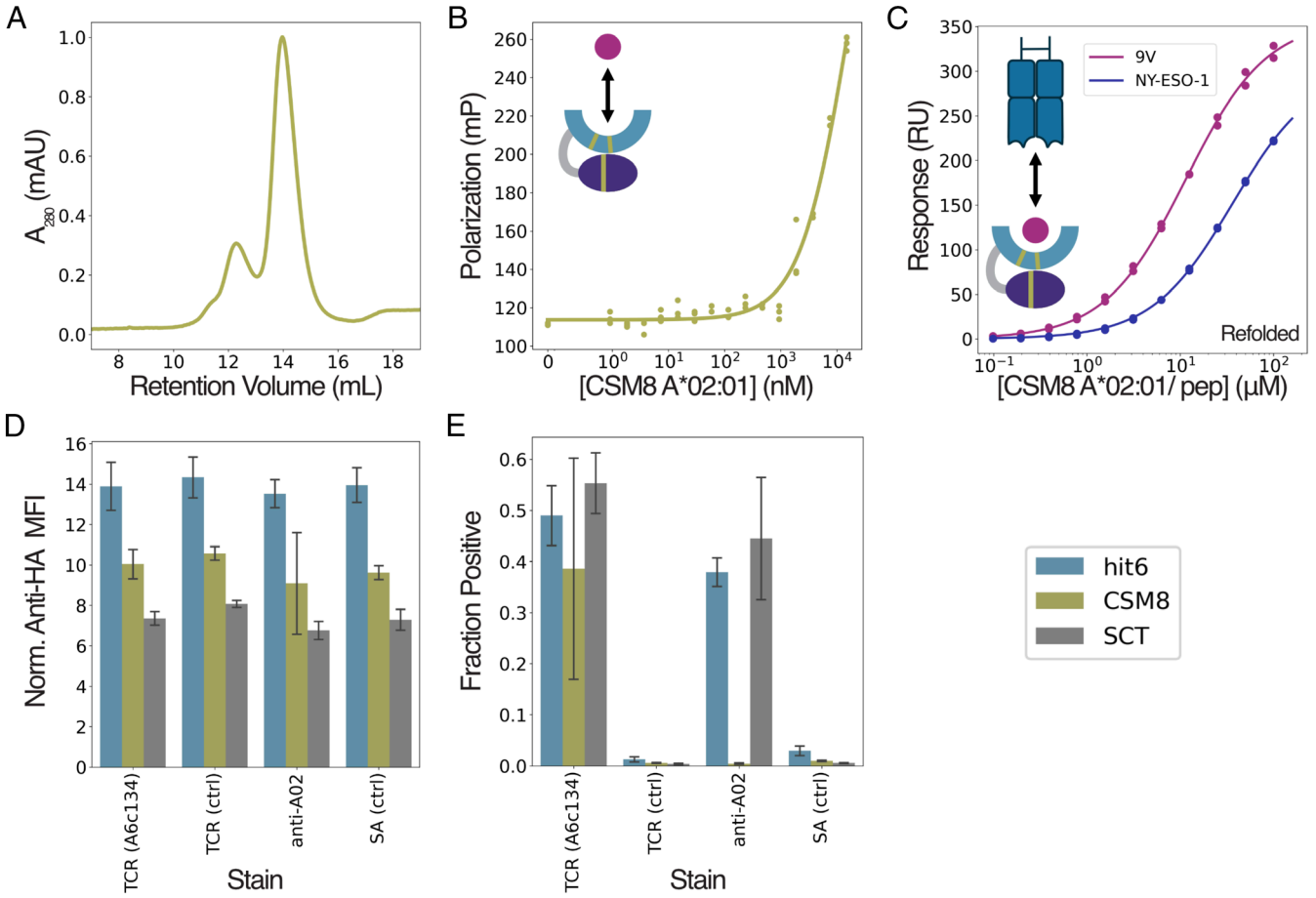

Fig. 3. CSM8 A*02:01 retains native TCR interactions and improves pMHC yeast display. (*A*) SEC trace of empty CSM8 A*02:01 purified from the soluble fraction of *E. coli* and thereafter from the monomeric fraction by SEC on an S200 column. (*B*) FP data (filled circles) and fitted binding curve (line) for $n = 3$ technical replicates of CSM8 A*02:01 binding to the AF488-NY-ESO-1 peptide. (*C*) Equilibrium binding data (filled circles) and fitted binding curves (lines) from SPR experiments with refolded CSM8 A*02:01 presenting variants of the NY-ESO-1 peptide to immobilized 1G4 TCR. (*D* and *E*) yeast display data for hit6 (teal), CSM8 (yellow), or full-length SCT (gray) A*02:01 stained with an anti-HA antibody and either SA A6c134 TCR tetramers, control TCR (TCR156) tetramers, anti-A*02 antibody, or a SA control. (*D*) Anti-HA MFIs for the HA+ population of each sample, normalized to the maximal MFI of the HA- population. (*E*) Fraction of HA+ yeast cells that are also stain+ for the anti-A*02 or TCR tetramer stain in each sample.

designs as CSM8-L11 throughout the remainder of the text (denoted by purple coloring). As intended, the rigid linker significantly improved yields in both *E. coli* and CFPS expression systems and reduced aggregation for H-2D^b without impacting the gp33 binding affinity (*SI Appendix*, Fig. S7 and Table S7).

We also found CSM8-L11 H-2D^b to be highly shelf-stable, retaining strong peptide binding affinity after storage for at least 1 mo at 4 °C, and showing only a minor decrease in affinity following multiple freeze/thaw cycles (*SI Appendix*, Fig. S8 *A* and *B*). To further assess peptide binding and stability, we measured circular dichroism (CD) spectra and melting curves for CSM8-L11 H-2D^b in the presence and absence of the gp33 peptide. The CD spectra are consistent with the mixed αβ fold of the design model (*SI Appendix*, Fig. S8*C*), while the melting curves demonstrate a clear stabilization of the fold in the presence of the peptide (*SI Appendix*, Fig. S8*D*).

After confirming that the rigid linker improved solubility and stability of H-2D^b, we tested its ability to improve these properties in A*02:01. We found that the yields were comparable (*SI Appendix*, Table S7), although CSM8-L11 A*02:01 was somewhat more prone to aggregation than CSM8 A*02:01 (*SI Appendix*, Fig. S9*A*). Despite the increased aggregation, CSM8-L11 showed a slight improvement in $K_{D,\,app}$ when binding to the NY-ESO-1 peptide, and similar A6c234 TCR binding in yeast display experiments (*SI Appendix*, Fig. S9 *B*–*E*). In combination with the improvements observed in H-2D^b, these results suggest that the rigid linker represents an overall improvement to SMART MHCs and therefore focused on this improved design in subsequent experiments.

Table 2. 1G4 TCR binding affinities for CSM8 A*02:01 in complex with NY-ESO-1 peptide variants

Peptide name	Peptide sequence	Soluble CSM8 A*02:01 affinity (μM)*	Refolded CSM8 A*02:01 affinity (μM)*	Native A*02:01 affinity (μM)†
NY-ESO-1	SLLM-WITQC	186 ± 16	36.5 ± 0.60	13.3 ± 0.4
NY-ESO-1-9V	SLLM-WITQ**V**	98.2 ± 4.2	11.5 ± 0.34	5.7 ± 0.2

CSM8 affinities are calculated by fitting a standard binding curve to the equilibrium binding SPR values. Mutations relative to the NY-ESO-1 sequence are shown in bold.
*Values are reported as fitted K_D ± 1 SD (as determined curve-fitting variance).
†Measured previously by Chen et al. (49).

CSM8-L11 Solubilizes Several Additional Common HLA Allomorphs

Given the improvements in expression observed with CSM8-L11, we expressed 15 human HLA allomorphs using this stabilizing domain in *E. coli* and performed small-scale purification using

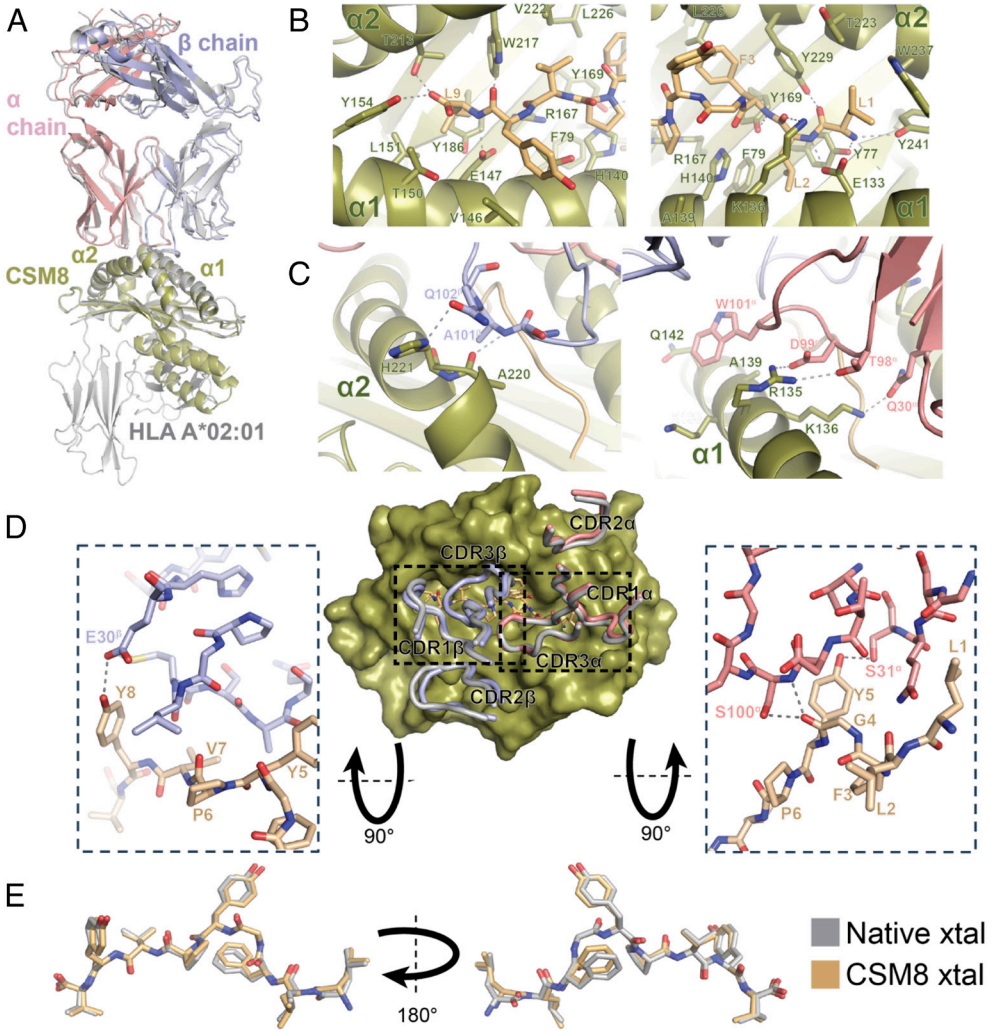

Fig. 4. The crystal structure of the CSM8 A*02:01/TAX9/A6c234 complex reveals a native-like TCR docking footprint. (*A*) crystal structure aligned to the HLA chain of the native A*02:01-TAX/A6c134 structure (PDB ID 4FTV, gray). For the CSM8 structure, the α chain is shown in salmon, the β chain in light blue, TAX9 peptide in light orange, and CSM8 A*02:01 in light green. The α1 and α2 helices of the CSM8 molecule are labeled for reference. (*B*) CSM8 A*02:01 interactions with the N terminus (*Left*) and C-terminus (*Right*) of TAX9 peptide. Interacting residues are shown as sticks, and hydrogen bonds are shown as dashed lines. (*C*) CSM8 A*02:01 interactions with the β (*Left*) and α (*Right*) chains of the A6c234 TCR. (*D*) *Top* view of CSM8 A*02:01 (shown as surface representation) with TCR CDR loops shown as tubes (*Center*); CDR loops from the native structure are shown in gray. Peptide contacts from the β chain (*Left*) and α chain (*Right*) are shown as *Insets*. (*E*) Comparison of the TAX9 peptide in the two crystal structures, aligned on the α1 and α2 domains of the presenting molecules.

HPLC. CSM8-L11 versions of many allomorphs exhibited peaks in the expected monomeric range, though many also displayed dimeric and aggregate peaks (Fig. 5*A*). Notably, HLA A*03:01 and HLA A*01:01 demonstrated high protein expression, and HLA B*07:02 showed the highest monomeric fraction (Fig. 5*B*). The presence of a significant dimeric or aggregated population, particularly in allomorphs with high expression, indicates that the stabilizer was ineffective at simultaneously promoting high protein expression and monomeric behavior across diverse HLAs. However, four allomorphs showed increased expression, and three showed reduced dimerization compared to CSM8-L11 A*02:01, suggesting potential for improvement through minor redesign.

We next compared the expression of native, full-length HLAs with their CSM8-L11 versions, with and without genetically linked peptides, using CFE. Soluble protein levels were measured using radiolabeled ^{14}C-Leucine incorporation (Fig. 5*C*). About 93% (27/29) of SMART HLA constructs showed increased soluble expression vs. their full-length counterparts, with an average 5.6-fold improvement and several exceeding 10-fold. Enhanced solubility in CFE may result from reduced crowding, a more oxidizing environment, or DsbC chaperone activity. Together, *E. coli* and CFE data suggest that while the stabilizing domain aids solubilization, further optimization is needed to reduce aggregation and enhance generalizability across HLAs.

Peptide-Fused SMART MHC Oligomers Stain T Cells in a TCR-Specific Manner

An important application of pMHCs is the staining and identification of T cells using pMHC tetramers (1). We therefore tested whether CSM8-L11 H-2D^b could be converted into a similar multimeric staining reagent. Rather than using SA to tetramerize our designs, as is typically done, we chose to directly fuse them to a de novo designed oligomeric protein which assembles into a tetrahedral architecture containing 12 subunits (57). This allowed us to omit the biotinylation step necessary for SA-based tetramerization (58). To improve folding and assembly of oligomeric CSM8-L11 H-2D^b, we fused a peptide of interest to the N terminus of the construct along with a SUMO tag. Coexpression of the Ulp1 protease allows the N terminus of the peptide to be cleanly cleaved, enabling it to bind properly in the peptide binding groove. The Y84A mutation in the H-2D^b sequence was used to accommodate the peptide linker (59). We fused a Myc tag to the C-terminus of the oligomer to allow for antibody staining (*SI Appendix*, Fig. S10 *A*–*C*).

To assess the ability of the CSM8-L11 H-2D^b MHC oligomers to stain T cells, we mixed two populations of Jurkat T cells: one expressing the P14 TCR (Fig. 6*A*) and the other expressing the unrelated TCR, OT-I (60) (*SI Appendix*, Fig. S10 *D* and *E*). To distinguish the two cell lines independently of TCR staining, we labeled the OT-I Jurkats with CFSE dye. We fused CSM8-L11

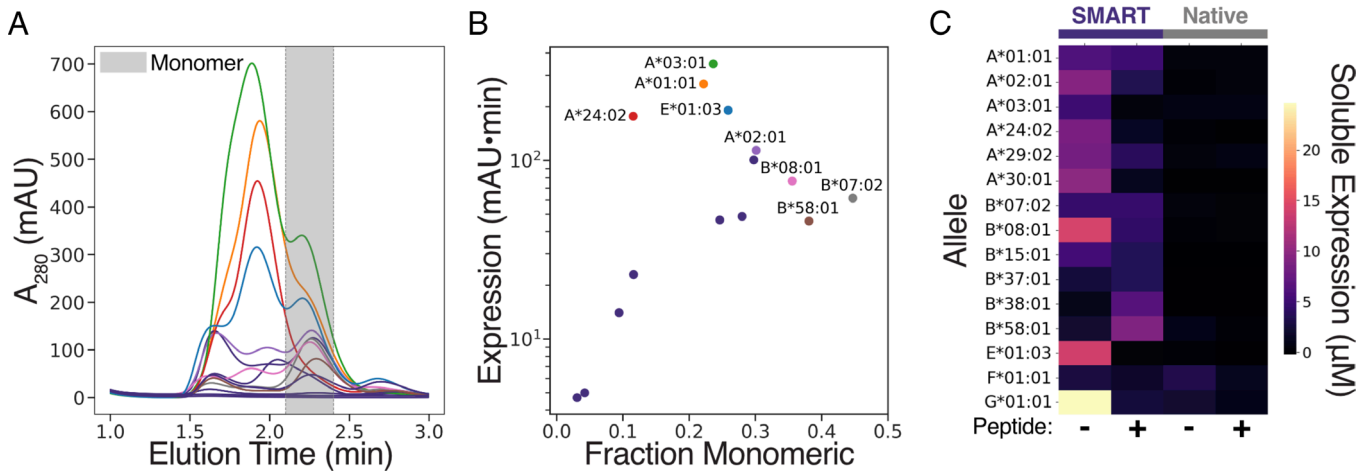

Fig. 5. Stabilization of multiple HLA alleles using the CSM8-L11 stabilizer. (*A*) SEC traces for empty CSM8-L11 HLAs. Expected monomer range for the S200 column on the HPLC is 2.1 to 2.4 min (shaded region). (*B*) Plot of CSM8-L11 HLA expression statistics showing total area under the SEC curve vs. the fraction of the sample that is monomeric. Colors for labeled points match colors in (*A*). (*C*) Heatmap of native and CSM8-L11 HLA CFE soluble expression levels with (+) and without (–) a linked peptide. Soluble protein (μM) was quantified using radioactive ^{14}C-leucine incorporation. Average of three replicates (n = 3) is shown for each construct.

H-2D^b oligomers to gp33 peptide variants with known affinities to the P14 TCR (29, 45) and assessed their ability to stain the T cell mixture. Our T cell staining data reflected previously observed trends in native pMHC/TCR binding (29, 45); the highest affinity variant (V3P) showed the brightest staining, with staining intensity decreasing with decreasing affinity (Fig. 6 and *SI Appendix*, Fig. S10). However, the Y4F variant showed clear staining in our experiments, despite prior work showing that the P14 TCR does not recognize this peptide (45). This non-native behavior matches our SPR measurements showing weak binding of the P14 TCR to CSM8 H-2D^b/Y4F (Fig. 2*C*), indicating that this behavior is consistent across experimental contexts. The detection of binding to Y4F in both experimental settings suggests that this difference from native behavior likely results from minor structural differences between SMART and native H-2D^b, rather than specific features of the oligomeric format used here. Nevertheless, staining was both TCR- and peptide-specific, suggesting that CSM8-L11 H-2D^b oligomers could be used alongside pMHC tetramers to more easily identify and isolate pMHC-specific T cells.

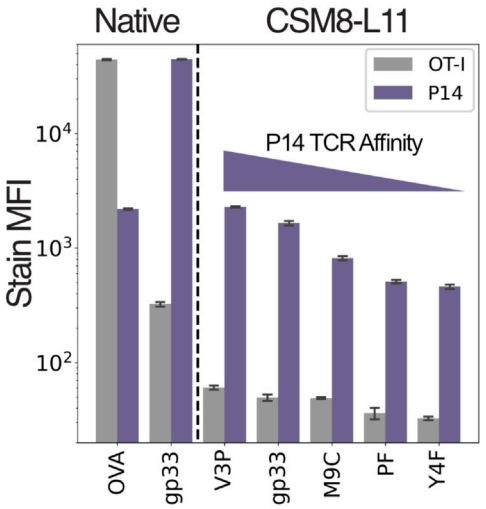

Fig. 6. Peptide-fused SMART MHC oligomers stain T cells in a TCR-specific manner. Bar plots indicating mean and SD (n = 3) of staining intensities for native tetramer controls (left of the dashed line) or CSM8-L11 oligomers fused to various gp33 peptide variants (right of the dashed line).

Discussion

Our success in creating a single stabilizing domain that allows soluble expression in *E. coli* of H-2D^b and partial solubilization of A*02:01 and other HLAs provides a strong foundation for a generalizable stabilizing scaffold for other MHC-I allomorphs. Our structural and biochemical data confirm that SMART H-2D^b maintains critical peptide and TCR interactions, while SMART A*02:01 shows weakened binding interactions. Our structural data also show that, for both allomorphs, the conformation of the presented peptide is nearly identical to the native conformation, and that for refolded CSM8 A*02:01, the TCR binds through the same set of interactions in both the native and SMART complexes. Recent work has also demonstrated that the reduced size of refolded SMART A*02:01 facilitates rapid characterization of peptide- and TCR- MHC interactions via NMR (61). Thus, SMART MHCs can provide an accessible complement to conventional MHCs, bypassing the need for refolding and improving stability for H-2D^b, and providing a convenient miniaturized form for A*02:01.

In addition to their utility in biophysical and T cell staining experiments, SMART MHCs could improve high-throughput measurements of pMHC/TCR interactions. Their high expression levels on the yeast surface could expedite screening of pMHC/TCR interactions in yeast display experiments. Additionally, the soluble expression of SMART MHCs in cell-free systems opens unique routes to characterize pMHC–TCR interactions in high throughput by taking advantage of the scalable nature of CFE (33, 37, 62, 63). Finally, the ability to produce a tetramer-like T cell staining reagent without labor-intensive refolding protocols has the potential to reduce the barriers to performing T cell tracking or sorting experiments. The SMART system could dramatically increase the number of peptide and MHC variants that can be tested, enabling improved tracking of immune responses to infectious disease, identification of cancer-targeting T cell clones, and many other applications. To fully realize this potential, the SMART design will need further improvement, as many HLAs are not expressible in functional form with the SMART domain fusion (Fig. 5). With such improvements, SMART MHCs have the potential to rapidly accelerate our understanding of T cell behavior and TCR specificity.

Limitations and Future Work. The SMART MHCs presented here are prototypes for a de-novo design strategy to produce small, modular MHC proteins; further refinement of these designs will be necessary before they can be used as general reagents with broad utility across MHC allotypes. While we were able to express a significant soluble fraction of H-2D^b in the cytosol of *E. coli*, our results with A*02:01 and other HLAs were less successful. Furthermore, only a fraction of soluble A*02:01 is in a peptide-receptive state, and this allotype was prone to dimerization and showed weakened binding affinities. While SMART A*02:01 showed display on yeast by TCR staining, library generation for TCR selection with this scaffold has not yet been achieved. Other SMART HLAs showed signs of similar liabilities, though they were not tested extensively. To overcome these limitations and make SMART MHCs broadly useful in a variety of biochemical and immunological applications, next generation versions are focused on screening for stabilizers that promote both peptide and TCR binding across diverse MHC allotypes.

Methods

Stabilizer Library Design. We used previously developed computational methods to design protein binders for arbitrary target proteins (23) 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). Briefly, we docked scaffolds to the underside of the truncated structure and designed favorable interactions, including some taken from the native interactions with β2m. Designs were filtered for the quality of the stabilizer and the interactions it made with the MHC, as previously described (23). We also included a set of negative control designs made by randomly scrambling the sequence (while conserving the pattern of hydrophobic and hydrophilic residues) of a random subset of the designs. A detailed description of this protocol is provided in https://github.com/wlwhite-tufts/code-for-SMART-MHC-manuscript (64).

Yeast Display Screening of Stabilizing Domains. Yeast display was performed as previously described (23) 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 side chain. Sorted populations were subsequently cultured, and plasmid DNA was extracted for sequencing. FITC-gp33 was purchased from GenScript as a custom peptide synthesis.

CFE and Radioactive Quantification of Soluble Protein Yields. CFE reactions were prepared with a version of the existing PANOx-SP formulation that uses glutathione and DsbC to create an oxidizing environment (34–36), and ^{14}C-leucine for later measurements of yield. All components were added together on ice (37, 39). This final reaction mixture was added to 6.66% v/v unpurified linear DNA PCR products (LETs) in triplicate (63, 65) and incubated at 30 °C overnight. Samples were spun at 16,000 × g for 10 min at 4 °C to separate total and soluble protein fractions. Supernatants were incubated at 37 °C for 20 min with 0.25 N KOH. Fractions were spotted on a 96-well filtermat (Revvity 1450–421), dried, washed, and coated with wax, and protein yields were quantified with a scintillation counter. Further details are provided in *SI Appendix, Supplemental Methods*.

Cleavage Site Mutation Design. Cleavage site design was used to make mutations to the MHC sequence to reduce proteolysis without impacting peptide or TCR binding. First, a multiple sequence alignment (MSA) of MHC protein sequences was collected using PSI-BLAST (66). Next, the MSA was converted into a position-specific score matrix (PSSM) denoting the likelihood of observing each possible amino acid at each position in the MSA. Standard Rosetta design protocols (67) were modified to allow mutations to amino acids that had likelihoods above a specified cutoff and to only allow mutations at sequence positions near the preidentified cleavage sites. Design was further restricted to prevent mutations in residues with side chains that could interact with a bound peptide or TCR. All CSMs were designed based on the hit6 design model, and the highest scoring designs were selected for experimental testing based on a combination of Rosetta metrics relating to the quality of the design model and strength of the interface. A detailed description of this protocol is provided in https://github.com/wlwhite-tufts/code-for-SMART-MHC-manuscript (64).

Peptide Binding Affinity Measurements. Three technical replicates of varying concentrations of SMART MHC were mixed with a constant concentration of fluorophore-labeled peptide (300 pM FITC-gp33 for H-2D^b and 10 nM AF488-NY-ESO-1 for A*02:01) and incubated overnight at room temperature to allow equilibration. FP measurements of these samples were made with a Synergy Neo2 plate reader (BioTek instruments) with a 485/530 FP filter. Binding curves [using the nonsimplified equilibrium binding equation (68)] were fitted separately to each of the triplicate measurements and averaged to determine the K_D. The peptides used in these experiments were synthesized in-house using the methods described above.

CD Measurements. Purified soluble SMART H-2D^b was diluted to 0.3 mg/mL (9.0 μM) in 25 mM Tris (pH 8.0), 150 mM NaCl, and 5% glycerol. The gp33 peptide stock was prepared by dissolving dry peptide (purchased from GenScript as a custom synthesis) to 2 mg/mL (1.52 mM) in methanol. Twofold molar excess of peptide stock (or an equivalent volume of methanol) was added to the protein sample and incubated at 4 °C overnight. CD spectra were measured using a Jasco J-1500, and melting curves were measured in increments of 0.5 °C at a rate of 2 °C/min.

SPR Measurements. TCR binding affinities were measured as previously described (45), using CSM8 H-2D^b as the mobile phase. All measurements were performed on a BIAcore T200 (GE Healthcare) at 20 °C in HBS with 0.005% Tween-20 and 3 mM EDTA. Soluble P14-6xHis (0.75 μM) was bound to immobilized anti-His on a CM5-chip. Varied concentrations of freshly produced CSM8 H-2D^b/peptide complexes were injected over the chip surfaces, maintained at 4 °C. Chip surfaces were regenerated using a low pH buffer after each injection. The final signal was calculated by subtracting the signal obtained on the control (no TCR) surface from the signal on the TCR-coupling surface. SPR data were analyzed with BIAevaluation 3.0 software (Cytiva), K_D values were obtained from steady–state fitting of equilibrium binding curves from at least 10 sample injections. Both A*02:01 and H-2D^b SPR measurements were performed using the same approach. Further details are provided in *SI Appendix, Supplemental Methods*.

CSM8 H-2D^b Crystallography. Crystallization of CSM8 H-2D^b/gp33 was performed using the sitting-drop vapor diffusion method at 293.15 K. Drops were set up using 0.15 μL CSM8 H-2D^b/gp33 and 0.15 μL 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 reservoir solution. Crystals appeared after 6 to 13 d and were cryoprotected with a solution containing an additional 7.5% w/v PEG 4,000 and harvested using mounted CryoLoops (Hampton Research). Subsequently, the crystals were flash-frozen in liquid nitrogen for transportation and 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 (69). The crystal structure was determined by molecular replacement with Phaser-MR in PHENIX (70) with the design model employed as the search model. Refinement was performed with PHENIX, followed by manual model building with Coot (71), and final refinement with PHENIX (72–74). Figures were generated using PyMOL Molecular Graphics System (Schrödinger). The final coordinates/structure factors have been deposited in the PDB with accession code 9HY4 (75).

Peptide-Fused Yeast Display. Yeast display of peptide-fused hit6, CSM8, and SMART A*02:01 was performed as previously described (19). In brief, 50 ng pCT or pYAL plasmids encoding corresponding full length or hit6, CSM8, and SMART A02 constructs with TAX were electroporated into competent EBY100. The EBY100 was cultured in YPD medium for 1 h at 30 °C, spun down, and continued to grow in SDCAA medium for 48 h before induction in SGCAA for 48 h. 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 SA.

A6c134/CSM8 A*02:01/TAX9 Crystallography. Purified A6c134 TCR was combined with equimolar refolded CSM8 A*02:01/TAX9 and purified on a Superdex S200 10/300 Increase column in HBS. Diffraction-quality crystals were grown by microseed matrix screening. Diffraction images were collected at NSLS-2

beamline 17-ID-1 at Brookhaven National Laboratory. The structure was solved by molecular replacement and refined using data to 3.4 Å resolution. Detailed crystallographic methods can be accessed in *SI Appendix*. An example of electron density at the final stage of refinement is shown in *SI Appendix*, Fig. S6. Data collection and refinement statistics are presented in *SI Appendix*, Table S2. Structure figures were generated using PyMOL. Structure factors and final model coordinates have been deposited in the PDB with accession code 9NDS, and diffraction images have been deposited in the SBGrid Databank.

Linker Design. Linker design was used to replace the flexible GS linker used in the yeast display screening with a shorter, structured linker. Starting with the design model of the CSM8 variant, we used existing "inpainting" methods (55) to fill in a small segment of protein structure, bridging the gap between the C-terminus of the stabilizer and N terminus of the MHC. Sequences for the resulting protein backbones were designed using ProteinMPNN (54), restricted to changing only the amino acids in the "inpainted" structure. Finally, the resulting designs were evaluated using AlphaFold2 (56) predictions with the MHC structure provided as a template. Designs with high overall pLDDT scores and low PAE scores for residues in the MHC/stabilizer interface were selected for experimental testing. A detailed description of this protocol is provided in https://github.com/wlwhite-tufts/code-for-SMART-MHC-manuscript (64).

Small-Scale Expression and SEC of HLA Alleles. Genes encoding CSM8-L11 HLAs were cloned into expression vector LM0627 (Addgene 191551) containing the MSG residues at the N terminus and a SNAC linked to 6xHis tag at the C-terminus, and transformed into *E. coli*. Cells were grown at 37 °C to an OD_{600} of 1 in TB-II with Kanamycin, induced with IPTG, and moved to 16 °C for overnight expression. Cells were pelleted and lysed with BugBuster®, and insoluble material was removed by centrifugation. Soluble lysate was purified in a 96-well plate using Ni-NTA affinity chromatography, filtered, and analyzed using HPLC (Agilent) with an S75 column (Cytiva).

T Cell Staining. OT-I TCR Jurkat cells were stained with CFSE and washed. OT-I and P14 Jurkat lines were mixed in equal numbers and resuspended to 1 M cells/mL in 100 μL of staining solution containing either a native MHC tetramer (PE/H-2K^b/OVA or APC/H-2D^b/gp33) or CSM8-L11 H-2D^b/peptide oligomer and incubated at 4 °C for 30 min, and washed. Oligomer-treated cells were then stained with AF647-anti-Myc, washed, and resuspended in HBH for analysis on an Attune Nxt flow cytometer. Further details are provided in *SI Appendix, Supplemental Methods*.

Data, Materials, and Software Availability. Atomic coordinates for the CSM8 H-2D^b/gp33 and CSM8 A*02:01/TAX9/A6c13 have been deposited in the PDB with the accession codes 9HY4 (75) and 9NDS (76) respectively. Data and code used to create the figures, as well as details of the computational design protocols, are provided on GitHub: https://github.com/wlwhite-tufts/code-for-SMART-MHC-manuscript (64). Sequences described in this paper are listed in *SI Appendix*, Tables S3–S6. All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. We thank members of the Kueh lab for discussion and feedback on the manuscript and Dr. Matthew Wither for providing the P14 and OT-I Jurkat cell lines. We also thank Dr. Brian Coventry and Dr. Lonxing Cao for the development of the protein design protocols used to create the yeast display library of stabilizing domains and for their help adapting their protocols to the context of MHC stabilization. We also thank Dr. Sangmin Lee for the development of the oligomer used in our T cell staining experiments. We also thank Dr. Robert Divine and Dr. David Feldman for the scripts used in reverse translation of the design library. We also thank Dr. Chris Norn and Dr. Kiera Sumida for the development of the protein design protocols used to create the CSMs and for their help running those protocols. Finally, we thank Lara Roché-Sudar for her help editing the manuscript. The Center for BioMolecular Structure is primarily supported by the NIH, National Institute of General Medical Sciences through Grant # P30GM133893, and by the Department of Energy (DOE) Office of Biological and Environmental Research FWP # BO070. This research used resources 17-ID-1 of the National Synchrotron Light Source II, a U.S. DOE Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. This work was funded by grants from the Bill and Melinda Gates Foundation (INV-010680; to D.B.), the NIH (5 R01 AI103867 08; to K.C.G.), the NIH/National Institute of Biomedical Imaging and Bioengineering Trailblazer Award (R21EB027327; to H.Y.K.), an NIH R01 (R01CA282512; to H.Y.K.), the Swedish Research Council (2021-05061; to A.A.), the Swedish Cancer Society (24 3775 Pj 01 H; to A.A.), the King Gustaf V Jubileum Fund (244092; to A.A.), the Swedish Cancer and Allergy Foundation (11338; to A.A.), the Yosemite fund, the Ludwig Institute for Cancer Research, the Defense Threat Reduction Agency (HDTRA1-21-1-0038; to M.C.J.), the Army Research Office Defense University Research Instrumentation Program (W911NF-23-1-0334; to M.C.J.), and a NSF Graduate Research Fellowship Program award (DGE-1656518; to L.G.). This work was delivered as part of the NexTGen team supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (CGCATF-2021/100002) and the National Cancer Institute (CA278687-01) and The Mark Foundation for Cancer Research. This work was delivered as part of the MATCHMAKERS team, supported by the Cancer Grand Challenges partnership funded by Cancer Research UK [CGCATF-2023/100004 (N.G.S.), CGCATF-2023/100006 (K.C.G.), and CGCATF-2023/100008 (D.B.)]; the National Cancer Institute [OT2CA297575 (N.G.S), OT2CA297242 (K.C.G.), 1-R33 CA302206-01 (K.C.G.), and OT2CA297288 (D.B.)]; and The Mark Foundation for Cancer Research (N.G.S, K.C.G., and D.B.). Portions of the text and figures were developed from the thesis of W.L.W.

Author affiliations: [a]Department of Bioengineering, University of Washington, Seattle, WA 98195; [b]Institute for Protein Design, School of Medicine, University of Washington, Seattle, WA 98195; [c]Department of Biochemistry, University of Washington, Seattle, WA 98105; [d]HHMI, University of Washington, Seattle, WA 98105; [e]Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA 94305; [f]Science for Life Laboratory, Department of Medicine Solna, Karolinska Institute and Division of Infectious Diseases, Karolinska University Hospital, Stockholm 171 77, Sweden; [g]Department of Bioengineering, Stanford University, Stanford, CA 94305; [h]Center for Computational and Genomic Medicine, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104; [i]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [j]HHMI, Stanford University School of Medicine, Stanford, CA 94305; [k]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA 98109; [l]Science for Life Laboratories, Karolinska Institute, Solna SE-171 65, Sweden; and [m]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305

Author contributions: W.L.W., H.B., C.J.K., K.M.J., R.S., L.G., X.H., X.C., J.E.B., Y.S., N.W., C.W., M.C.J., N.G.S., A.A., K.C.G., and D.B. designed research; W.L.W., H.B., C.J.K., K.M.J., R.S., L.G., X.H., X.C., A.C., J.E.B., Y.S., A.E.O., N.W., C.W., P.-Å.N., X.L., I.G., A.A., and P.M.L. performed research; X.L., P.M.L., and H.Y.K. contributed new reagents/analytic tools; W.L.W., C.J.K., K.M.J., R.S., L.G., X.H., X.C., J.E.B., Y.S., A.E.O., and X.L. analyzed data; H.Y.K. provided lab space and resources; and W.L.W., K.M.J., L.G., N.G.S., A.A., and D.B. wrote the paper.

Competing interest statement: Several authors (W.L.W., H.B., C.J.K., R.S., X.C., I.G., A.A., A.A., H.Y.K., K.C.G., and D.B.) are co-inventors on a provisional patent related to this work.

1. L. Wooldridge *et al.*, Tricks with tetramers: How to get the most from multimeric peptide–MHC. *Immunology* **126**, 147–164 (2009).
2. L. V. Sibener *et al.*, Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* **174**, 672–687.e27 (2018).
3. J. Pettmann, *et al.*, The discriminatory power of the T cell receptor. *eLife* **10**, e67092 (2021).
4. C. Linnemann *et al.*, High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat. Med.* **19**, 1534–1541 (2013).
5. K. Murali-Krishna *et al.*, Counting antigen-specific CD8 T cells: A reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
6. D. Kronenberg *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).
7. L. C. Rowntree, *et al.*, Preferential HLA-B27 allorecognition displayed by multiple cross-reactive antiviral CD8+ T cell receptors. *Front. Immunol.* **11**, 248 (2020).
8. S. Li, *et al.*, Characterization of neoantigen-specific T cells in cancer resistant to immune checkpoint therapies. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025570118 (2021).
9. C. J. Cohen *et al.*, Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J. Clin. Invest.* **125**, 3981–3991 (2015).
10. P. P. Lee *et al.*, Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat. Med.* **5**, 677–685 (1999).
11. P. Zareie *et al.*, Canonical T cell receptor docking on peptide-MHC is essential for T cell signaling. *Science* **372**, eabe9124 (2021).
12. L. Sušac *et al.*, Structure of a fully assembled tumor-specific T cell receptor ligated by pMHC. *Cell* **185**, 3201–3213.e19 (2022).

13. NIH Tetramer Core Facility, Production protocols. https://tetramer.yerkes.emory.edu/support/protocols#4 (2010).
14. A. H. Bakker *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. U.S.A.* **105**, 3825–3830 (2008).
15. J. J. Luimstra *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).
16. S. A. Overall *et al.*, High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat. Commun.* **11**, 1909 (2020).
17. Y. Sun *et al.*, Universal open MHC-I molecules for rapid peptide loading and enhanced complex stability across HLA allotypes. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2304055120 (2023).
18. M. E. Birnbaum, S. Dong, K. C. Garcia, Diversity-oriented approaches for interrogating T-cell receptor repertoire, ligand recognition, and function. *Immunol. Rev.* **250**, 82–101 (2012).
19. J. J. Adams *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).
20. F. Crawford, E. Huseby, J. White, P. Marrack, J. W. Kappler, Mimotopes for alloreactive and conventional T cells in a peptide–MHC display library. *PLoS Biol.* **2**, e90 (2004).
21. C. S. Dobson *et al.*, Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **19**, 449–460 (2022).
22. X. Yang *et al.*, Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature* **612**, 771–777 (2022).
23. L. Cao *et al.*, Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
24. D. R. Madden, The three-dimensional structure of peptide–MHC complexes. *Annu. Rev. Immunol.* **13**, 587–622 (1995).
25. R. Wang, K. Natarajan, D. H. Margulies, Structural basis of the CD8αβ/MHC class I interaction: Focused recognition orients CD8β to a T cell proximal position12. *J. Immunol.* **183**, 2554–2564 (2009).
26. L. L. Jones *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).
27. M. E. Birnbaum *et al.*, Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
28. E. T. Boder, K. D. Wittrup, Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
29. J. M. Boulter *et al.*, Potent T cell agonism mediated by a very rapid TCR/pMHC interaction. *Eur. J. Immunol.* **37**, 798–806 (2007).
30. A. Achour *et al.*, A structural basis for LCMV immune evasion: Subversion of H-2Db and H-2Kb presentation of gp33 revealed by comparative crystal structure analyses. *Immunity* **17**, 757–768 (2002).
31. A. Achour *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).
32. Z. Li *et al.*, The mechanism of β2m molecule-induced changes in the peptide presentation profile in a bony fish. *iScience* **23**, 101119 (2020).
33. A. D. Silverman, A. S. Karim, M. C. Jewett, Cell-free gene expression: An expanded repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).
34. A. D. Silverman, N. Kelley-Loughnane, J. B. Lucks, M. C. Jewett, Deconstructing cell-free extract preparation for in vitro activation of transcriptional genetic circuitry. *ACS Synth. Biol.* **8**, 403–414 (2019).
35. Y.-C. Kwon, M. C. Jewett, High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **5**, 8663 (2015).
36. M. C. Jewett, J. R. Swartz, Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* **86**, 19–26 (2004).
37. A. C. Hunt *et al.*, A rapid cell-free expression and screening platform for antibody discovery. *Nat. Commun.* **14**, 3897 (2023).
38. E. D. Carlson, R. Gan, C. E. Hodgman, M. C. Jewett, Cell-free protein synthesis: Applications come of age. *Biotechnol. Adv.* **30**, 1185–1194 (2012).
39. J. C. Stark *et al.*, Rapid biosynthesis of glycoprotein therapeutics and vaccines from freeze-dried bacterial cell lysates. *Nat. Protoc.* **18**, 2374–2398 (2023).
40. C. S. Sevier, C. A. Kaiser, Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **3**, 836–847 (2002).
41. M. Stech, S. Kubick, Cell-free synthesis meets antibody production: A review. *Antibodies* **4**, 12–33 (2015).
42. T. B. Dick, "Assembly of MHC class I peptide complexes from the perspective of disulfide bond formation." *Cell. Mol. Life Sci.* **61**, 547–556 (2004).
43. R. Vita *et al.*, The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
44. M. T. Puglielli *et al.*, In vivo selection of a lymphocytic choriomeningitis virus variant that affects recognition of the GP33-43 epitope by H-2Db but not H-2Kb. *J. Virol.* **75**, 5099–5107 (2001).
45. A. D. Duru *et al.*, Tuning antiviral CD8 T-cell response via proline-altered peptide ligand vaccination. *PLoS Pathog.* **16**, e1008244 (2020).
46. L. M. Velloso, J. Michaëlsson, H.-G. Ljunggren, G. Schneider, A. Achour, Determination of structural principles underlying three different modes of lymphocytic choriomeningitis virus escape from CTL recognition. *J. Immunol.* **172**, 5504–5511 (2004).
47. F. F. Gonzalez-Galarza *et al.*, Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2020).
48. U. Utz, D. Banks, S. Jacobson, W. E. Biddison, 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).
49. J.-L. Chen *et al.*, Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J. Exp. Med.* **201**, 1243–1255 (2005).
50. Y. T. Chen *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).
51. A. C. McShan *et al.*, Tapbpr promotes antigen loading on MHC-I molecules using a peptide trap. *Nat. Commun.* **12**, 3174 (2021).
52. A. Sachs *et al.*, "Impact of cysteine residues on MHC binding predictions and recognition by tumor-reactive T cells." *J Immunol.* **205**, 539–549 (2020).
53. D. K. Cole, Increased peptide contacts govern high affinity binding of a modified TCR whilst maintaining a native pMHC docking mode. *Front. Immunol.* **4**, 168 (2013).
54. J. Dauparas *et al.*, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
55. J. Wang *et al.*, Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
56. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
57. S. Lee *et al.*, Four-component protein nanocages designed by programmed symmetry breaking. *Nature* **638**, 546–552 (2025).
58. J. D. Altman *et al.*, Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
59. L. Lybarger *et al.*, Enhanced immune presentation of a single-chain major histocompatibility complex class I molecule engineered to optimize linkage of a C-terminally extended peptide. *J. Biol. Chem.* **278**, 27105–27111 (2003).
60. K. A. Hogquist *et al.*, T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
61. C. H. Woodward *et al.*, Solution mapping of MHC-I:TCR interactions using a minimalistic protein system. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2506016122 (2025).
62. A. C. Hunt *et al.*, Cell-free gene expression: Methods and applications. *Chem. Rev.* **125**, 91–149 (2025).
63. A. C. Hunt *et al.*, Multivalent designed proteins neutralize SARS-CoV-2 variants of concern and confer protection against infection in mice. *Sci. Transl. Med.* **14**, eabn1252 (2022).
64. W. L. White *et al.*, code-for-SMART-MHC-manuscript. Zenodo. https://doi.org/10.5281/zenodo.17966987. Deposited 17 December 2025.
65. Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, R. M. Murray, Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. *ACS Synth. Biol.* **3**, 387–397 (2014).
66. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
67. S. J. Fleishman *et al.*, RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).
68. E. C. Hulme, M. A. Trevethick, Ligand binding assays at equilibrium: Validation and interpretation. *Br. J. Pharmacol.* **161**, 1219–1237 (2010).
69. C. Vonrhein *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
70. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
71. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
72. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
73. D. S. Berkholz, M. V. Shapovalov, R. L. Dunbrack, P. A. Karplus, Conformation dependence of backbone geometry in proteins. *Structure* **1993**, 1316–1325 (2009).
74. C. J. Williams *et al.*, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
75. R. Sun *et al.*, Solubly expressed miniaturized SMART H2-Db. Protein Data Bank. https://doi.org/10.2210/pdb9hy4/pdb. Deposited 9 January 2025.
76. K. M. Jude *et al.*, Miniaturized HLA A*02-TAX bound to A6c134 TCR. Protein Data Bank. https://doi.org/10.2210/pdb9nds/pdb. Deposited 18 February 2025.