Yeast display platform with expression of linear peptide epitopes for high-throughput assessment of peptide-MHC-II binding

Received for publication, August 10, 2022, and in revised form, January 6, 2023 Published, Papers in Press, January 14, 2023,
https://doi.org/10.1016/j.jbc.2023.102913

Brooke D. Huisman[1,2], **Pallavi A. Balivada**[1,2], **and Michael E. Birnbaum**[1,2,3,*]

From the [1]*Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, USA;* [2]*Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA;* [3]*Ragon Institute of MGH, MIT and Harvard, Cambridge, USA*

Edited by Peter Cresswell

Yeast display can serve as a powerful tool to assess the binding of peptides to the major histocompatibility complex (pMHC) and pMHC-T-cell receptor binding. However, this approach is often limited by the need to optimize MHC proteins for yeast surface expression, which can be laborious and may not yield productive results. Here we present a second-generation yeast display platform for class II MHC molecules (MHC-II), which decouples MHC-II expression from yeast-expressed peptides, referred to as "peptide display." Peptide display obviates the need for yeast-specific MHC optimizations and increases the scale of MHC-II alleles available for use in yeast display screens. Because MHC identity is separated from the peptide library, a further benefit of this platform is the ability to assess a single library of peptides against any MHC-II. We demonstrate the utility of the peptide display platform across MHC-II proteins, screening HLA-DR, HLA-DP, and HLA-DQ alleles. We further explore parameters of selections, including reagent dependencies, MHC avidity, and use of competitor peptides. In summary, this approach presents an advance in the throughput and accessibility of screening peptide-MHC-II binding.

Yeast-displayed peptide–major histocompatibility complex (MHC) constructs, originally developed for assessing T-cell receptor (TCR) binding (1–3), have also proven useful for generating large, high-quality datasets on peptide-MHC binding (4–7). In previously described peptide-class II MHC (pMHC-II) yeast display approaches, the peptide is linked to the N terminus of MHC β chain, which is then expressed in either a β1α1 "mini" (1, 8) or β1β2/α1α2 "full length" format (3, 4, 6, 7). MHC-II proteins have also been expressed on yeast without linked peptide or covalent linkage of MHC α and β, utilizing leucine zipper fusions to promote heterodimer formation and assessing binding of exogenously added, chemically synthesized peptides (5). An additional approach expresses a secreted MHC-II in yeast for codisplay with a yeast surface-expressed peptide (9). While theoretically any MHC allele can be analyzed by yeast display, the relatively simple protein folding machinery in yeast can cause some MHCs, especially MHC-IIs, to not fold properly despite detectable surface expression (1, 4, 8). This necessitates each MHC be validated for folding, such as through the binding of a recombinantly expressed TCR (1, 8). In the absence of detectable TCR binding, error-prone mutagenesis of the MHC is required to identify mutations that can stabilize the MHC fold while not affecting the peptide- and receptor-binding surfaces (1, 8). This process can be time and effort intensive, has uncertain success, and requires availability of reagents such as a known MHC-restricted TCR.

To circumvent the need for individual optimization of MHC proteins, we have developed a second-generation yeast display approach for assessing pMHC-II binding, which does not require yeast-specific MHC-II optimizations. In this approach, peptide and MHC-II expression are decoupled, with a library of peptides displayed on the surface of yeast and MHC-II protein expressed as recombinant soluble molecules. Because of this decoupling and exclusive presentation of peptides on the yeast surface, we refer to this approach as "peptide display."

Here, we outline the system, explore experimental conditions, and compare the results with existing yeast display datasets. We apply this approach to study HLA-DR alleles, followed by HLA-DP and HLA-DQ alleles, which have to this point been largely inaccessible in previous yeast display approaches due to few HLA-DP or HLA-DQ alleles being validated as folded on the yeast surface (5). Existing pMHC yeast display platforms further require generation of a new library for each new MHC being screened (1, 2, 4, 6–8). In contrast, because the peptide library is decoupled from MHC, a single peptide display library is extensible to any MHC-II without requiring additional library cloning and generation, expanding the utility of the approach.

Results

Design of peptide display libraries and experiments

For the peptide display approach, we adapted the pCT302 yeast display vector, which fuses peptide sequences of interest to the C terminus of the yeast mating protein Aga2 (10). We left the N terminus of the construct unchanged, which

* For correspondence: Michael E. Birnbaum, mbirnb@mit.edu.

contains Aga2 fused to an epitope tag and a subsequent C-terminal linker sequence (Fig. 1*A*). To the C terminus of this linker, we connected the N terminus of a peptide sequence of interest. In order to readily assess frameshift mutations in our peptide sequence, we linked the displayed peptide's C terminus to a Myc epitope tag *via* a short linker.

To examine MHC-peptide repertoires, MHC-II proteins were expressed recombinantly rather than coexpressing MHC-II proteins on the surface of yeast, which circumvents yeast protein folding concerns. MHC-II proteins were stabilized with a linked portion of class II–associated Ii peptide ($CLIP_{87\text{-}101}$, PVSKMRMATPLLMQA) (11), connected to the

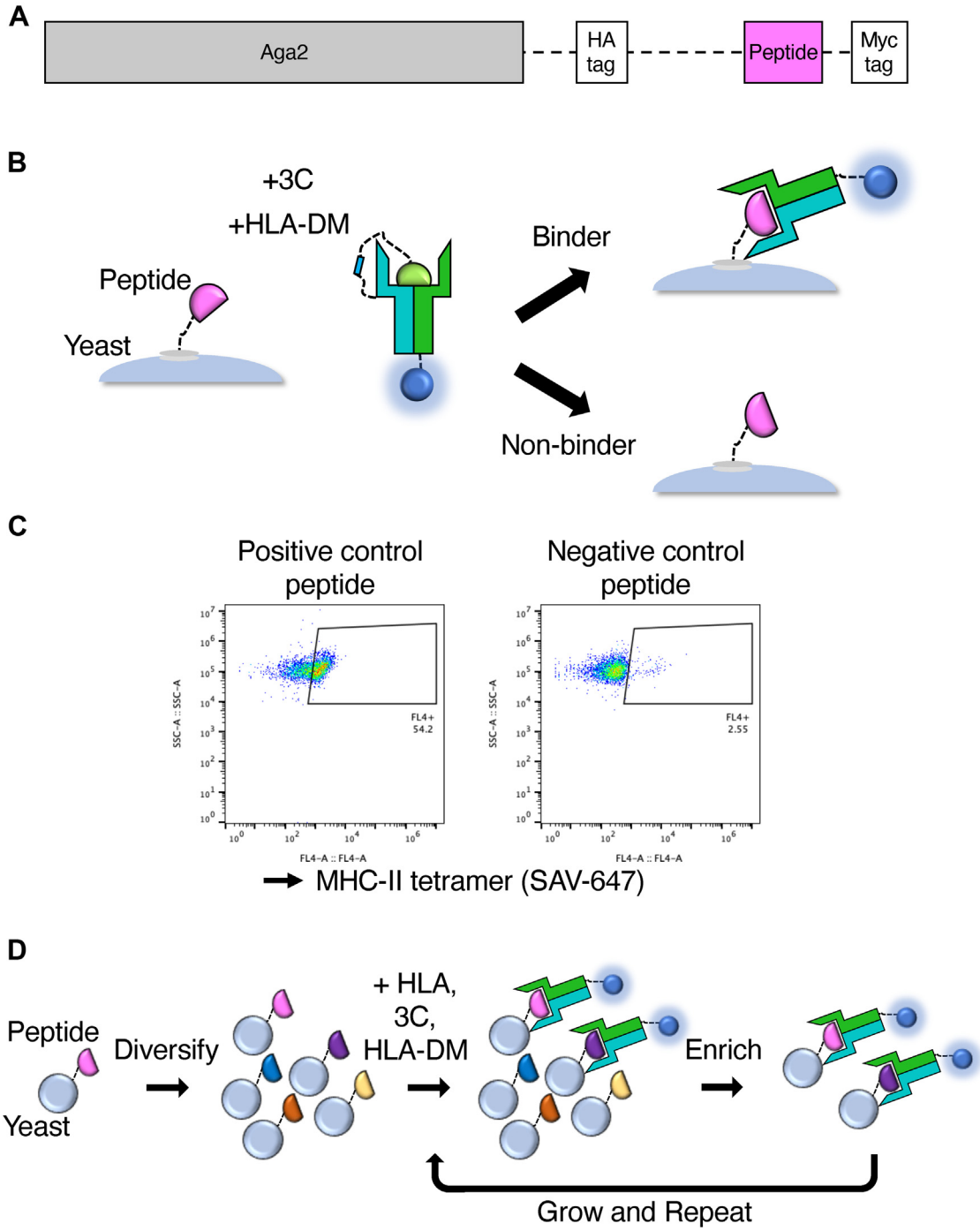

Figure 1. Overview of peptide display formatting and experiments. *A*, peptide display region of the pCT302 construct. A peptide and epitope tags are expressed on the surface of yeast, fused to the C terminus of the yeast mating protein Aga2. *B*, schematic of the peptide display platform formatting and selection strategy. Peptides (*pink*) are expressed on the surface of yeast *via* Aga2 (*gray*), and MHC is expressed recombinantly, with CLIP peptide (*green*) for stability. Upon addition of HLA-DM and 3C protease to cleave the CLIP-MHC linkage, the MHC can bind to yeast-expressed peptides. Binders can be manipulated using a handle on the MHC, such as a fluorescent streptavidin. *C*, flow cytometric analysis of tetramerized HLA-DR401 with binder (HA-derived) and nonbinder (CD48-derived) peptides. *D*, schematic of the library selection strategy, in which MHC-binding peptides are enriched in sequential rounds of selection.

N terminus of MHC-IIβ chain *via* a linker containing a 3C protease cleavage site. The MHC-II protein was site-specifically biotinylated, allowing for easy functionalization and visualization with fluorescent streptavidin (SAV).

To assess pMHC-II binding, yeast were coincubated with fluorescent tetramerized MHC-II protein, 3C protease, and HLA-DM (Fig. 1*B*). 3C protease cleaves the CLIP-MHC linker, enabling the stabilizing peptide to dissociate (Fig. S1). The peptide exchange catalyst HLA-DM is available to assist with dissociation of CLIP and binding to the yeast-expressed peptide (12). If the MHC is able to bind to the peptide of interest, we can detect fluorescence *via* flow cytometry–based assessment of the yeast population, akin to a primary antibody stain. When we stain peptide-displaying yeast with soluble tetramerized HLA-DR401, we observe a clear fluorescent population for yeast expressing a version of an Influenza A $HA_{306-318}$ peptide (APKYVKQNTLKLAT) known to bind HLA-DR401 (HLA-DRA1*01:01/DRB1*04:01) (4, 13), with minimal binding to an off-target $CD48_{36-51}$ peptide (FDQKIVEWDSRKSKYF) (Fig. 1*C*). The $CD48_{36-51}$ peptide binds to a related allele, HLA-DR402 (HLA-DRA1*01:01/DRB1*04:02) (4), but is likely dispreferred for HLA-DR401 binding because of the large aromatic Trp residue at the expected P4 pocket in the optimal register.

To explore the constraints of the system and reliance on each component of the selection reaction mix, we separately titrated HLA-DR401 tetramers, HLA-DM, and 3C protease and assessed pMHC-II binding. As expected, there was an MHC tetramer concentration dependence, with binding signal decreasing as MHC concentration decreased (Fig. S2*A*). We similarly observed a dependence on the presence of HLA-DM, with largely unchanged signal across concentrations from 0.5 μM to 3.5 μM, and with signal lost at the lowest concentration tested, 75 nM (Fig. S2*B*). Next, we assessed if the peptide exchange reaction could proceed without cleaving the linker between recombinant MHC-II and the stabilizing CLIP peptide. We observed loss of signal when we decreased 3C concentrations, illustrating a need to allow the stabilizing peptide to freely dissociate in order to enable loading of the yeast-expressed peptide (Fig. S2*C*). Across all experiments, the negative control peptide showed minimal signal at reagent concentrations tested. From these experiments, optimized reagent concentrations were selected for utilization in large-scale experiments (75 nM MHC-II tetramer, 0.5 μM HLA-DM, and 1 μM 3C).

Application of peptide display platform to screen HLA-DR401

Next, to utilize the peptide display approach for screening peptides at a repertoire scale, we generated a library of randomized 13mer peptides, containing approximately 5×10^6 unique peptides. We performed selections with HLA-DR401, as illustrated in Figure 1*D*, which would enable comparison with existing pMHC-II datasets (4, 14, 15). Yeast were incubated with fluorescent tetramerized HLA-DR401, 3C protease, and HLA-DM in phosphate-buffered saline (PBS) for 45 min. The neutral pH allows 3C protease to cleave the linker that connects MHC-II to the stabilizing CLIP peptide. To mimic acidic endosomal conditions, yeast were then incubated with the reaction mixture in pH 5 saline for an additional 45 min. Next, yeast were incubated with antifluorophore magnetic beads. Yeast that expressed peptides that bound to HLA-DR401 were enriched utilizing a magnetic enrichment strategy.

We performed three iterative rounds of selection, growing yeast to confluence between each round. In each round, we sampled yeast after the first incubation in PBS, the second incubation in acidic saline, and after elution from the magnetic column and examined the fluorescent population *via* flow cytometry. Over subsequent rounds, the population of fluorescent MHC-bound yeast increased at each of these steps (Fig. S3).

Next, we determined the identities of the peptides enriched for binding to HLA-DR401 by deep sequencing each round of the library selections. Given the open MHC-II binding groove and 13mer length of peptides assessed, we inferred the binding register of each peptide utilizing a position weight matrix approach (7). Peptides inferred to bind using the central register are represented in Figure 2. These representations include fractions of peptides with amino acids at each position (Fig. 2*A*), $\log_2$ fold change compared with the unselected library (Fig. 2*B*), and sequence logos (Fig. 2*C*). We observed a strong P1 preference for large hydrophobic and aromatic hydrophobic residues; a preference at P4 for acidic residues and nonaromatic hydrophobic residues; a preference for polar residues at P6; and preference for small residues at P9. The auxiliary anchor P7 exhibits some preference for Pro, Leu, and Asn. This motif is consistent with motifs from other high-throughput datasets (4, 14–17), with a greater representation of P4 hydrophobic residues relative to acidic residues, although peptides with hydrophobic P4 residues have been shown to bind to HLA-DR401 (4). Using the peptide display approach, we were able to enrich for peptides that bind to HLA-DR401, which are consistent with previously described motifs and examples of HLA-DR401-binding peptides.

We next assessed binding to a panel of individual peptides with known affinity values for HLA-DR401 (4). Performing binding reactions similar to selection experiments, we utilized flow cytometry to examine binding after the first incubation in PBS and the second incubation in acidic saline (Figs. 3 and S4). After the incubation in acidic saline, we observed a clear monotonic relationship between the measured IC_{50} value and percent MHC-tetramer positive population or mean fluorescence intensity. This relationship is only clear after the second incubation, which suggests that the acidic incubation may be helpful in enabling pMHC-II binding or establishing the protonation state of acidic residues. The clear binding of peptides containing acidic residues also suggests that acidic residue–containing peptides, including peptides with acidic P4 residues, are not systemically excluded from MHC binding in this format at an individual peptide level.

To explore the requirement for MHC avidity and to examine alternative modes of selections, we performed

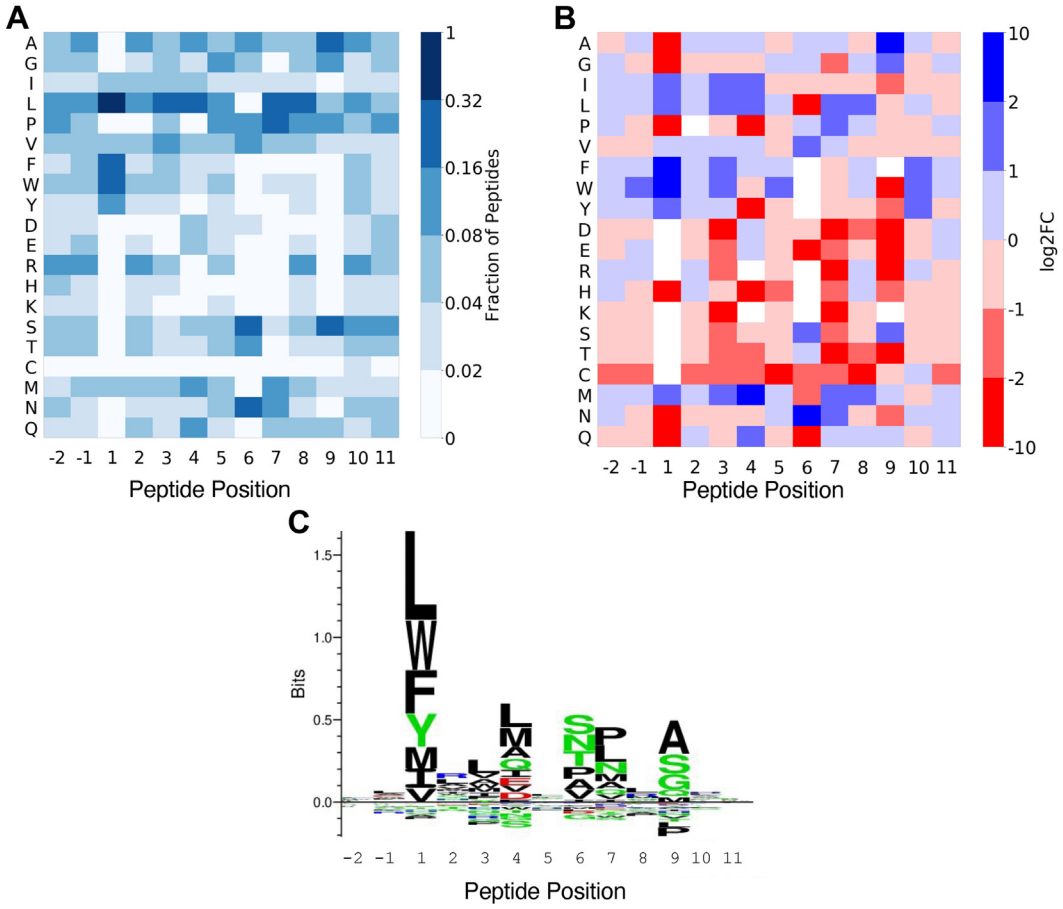

Figure 2. Motifs of enriched peptides in the central register from HLA-DR401 tetramer-based library selections. *A*, heatmap highlighting the fraction of peptides with each amino acid at each position. *B*, $\log_2$ fold change of positional amino acid frequency compared with unselected library. Amino acids that did not appear in the enriched sequences at a given position are *white*. *C*, sequence logo of enriched peptides.

selections with monomeric MHC. Performing monomeric selections also provides us a mode to assess if charged Alexa Fluor 647, which is conjugated to our streptavidin, could be affecting P4 acidic preference. Motifs from monomer selections (Fig. S5) are largely similar to the tetrameric selections (Fig. 2), suggesting MHC avidity in tetrameric selection is not necessary and the use of streptavidin–Alexa Fluor 647 (SAV-647) does not bias peptide binding.

Assessing additional MHCs, including HLA-DP and HLA-DQ alleles

Given the clear enrichment with HLA-DR401 we set out to demonstrate the extensibility of this approach to other MHC proteins by performing selections for HLA-DQ0602 (HLA-DQA1*01:02/DQB1*06:02), HLA-DQ0603 (HLA-DQA*01:02/DQB1*06:03), HLA-DP401 (HLA-DPA1*01:03/DPB1*04:01), and HLA-DR15 (HLA-DRA1*01:01/HLA-DRB1*15:01). These proteins represent alleles expressed by each of the three canonical MHC-II genes (18). Furthermore, these alleles are implicated in autoimmunity, with HLA-DQ0602 and HLA-DR15 comprising a haplotype linked to protection from type 1 diabetes, as well as susceptibility to multiple sclerosis (19–21). To ensure proper expression, we stabilized HLA-DP and HLA-DQ alleles with an extended $CLIP_{81\text{-}101}$ peptide (LPKPPKPVSKMRMATPLLMQA).

For all four alleles, we see clear enrichment, with preferences at MHC anchor residues (Fig. 4). HLA-DR15 prefers hydrophobic residues at P1; aromatic residues at P4; Asn, Ser, and Gly at P6; and small hydrophobic Leu, Val, and Ala at P9 (Fig. 4*A*). These preferences closely resemble previously reported motifs (14, 16). HLA-DP401 has strong preferences at P1, P6, and P9 for hydrophobic residues, including aromatics at P1 and P6. This motif is similar to those in existing datasets (14, 16). However, while somewhat present and most visible in the heatmaps (Fig. 4*B*), the preference for P4 Glu, Thr, and Ser is less pronounced than in the mass spectrometry comparison datasets (14, 16). HLA-DQ0602 has clear preference at P1, P4, P6, P7, and P9, alongside greater amino acid preference at nonanchors compared with the other alleles studied here (Fig. 4*C*). The anchor preferences and increased preference across the peptide is generally consistent with previously reported peptides from mass spectrometry (14). The peptide display approach can also elucidate differences between closely related alleles. A related HLA-DQ allele, HLA-DQ0603, showed similar P1, P4, and P9 preferences to HLA-DQ0602. The beta chain of HLA-DQ0603 differs from the HLA-DQ0602 beta chain at two amino acids, Phe9Tyr and

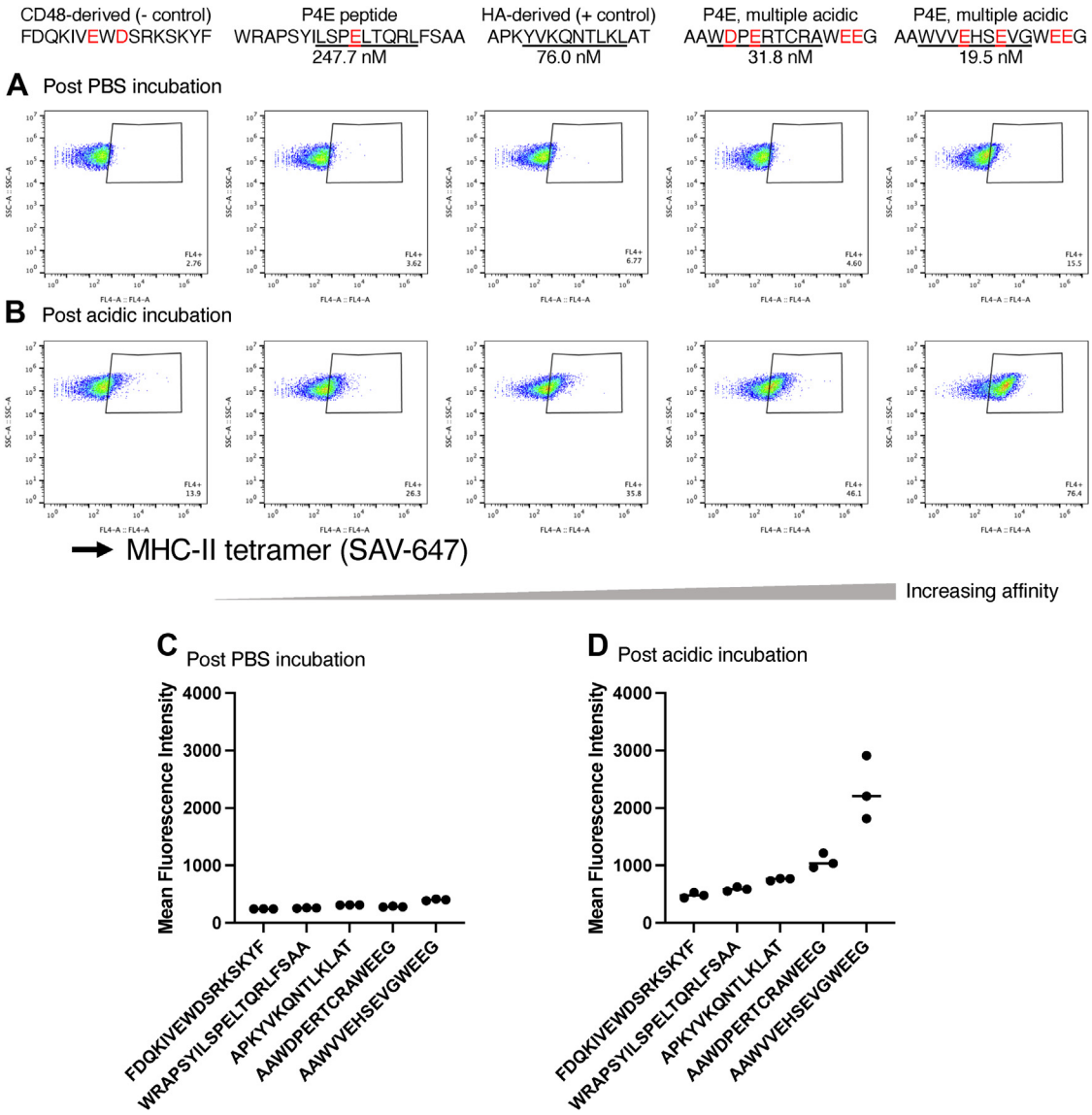

Figure 3. Tests on clonal yeast populations. HLA-DR401 tetramer staining on individual yeast clones after *A* incubation in pH 7.2 PBS and *B* subsequent incubation in acidic saline. CD48-derived peptide was included as a negative control, and the remaining peptides have previously measured affinities indicated. HA-derived peptide affinity was previously measured with an additional C-terminal Gly, which is present in the peptide display C-terminal linker. Central 9mer cores from inferred registers, based on similarity with HLA-DR401 peptide motifs, are underlined for each peptide, and acidic residues are highlighted in *red*. Data from replicate experiments are summarized in *C* and *D*.

Tyr30His, which affect P6 and P7 preferences (Figs. S6*A* and 4, *C* and *D*) (22). HLA-DQ0603 exhibits a greater preference for P6 Asp and Ser, likely due to the positive His at HLA-DQ beta position 30 (Fig. S6*B*, adapted from Protein Data Bank 6DIG (23)). Similarly, these polymorphisms generate a distinctive preference for Pro at the auxiliary P7 anchor, which is not captured in the NetMHCIIpan4.0 motif viewer (17). These data demonstrate the utility of the peptide display approach for extending yeast display assessment of peptide-MHC binding across alleles, especially alleles not previously validated on yeast.

Adding exogenous peptide as a competitor

The peptide display approach is highly tunable across multiple parameters. As an example, one such tunable parameter is the addition of competitor peptide. To examine the effects of adding exogenous peptide on MHC binding to yeast-expressed peptides, we titrated competitor peptide and assessed its effects on peptide-MHC-II binding (Figs. 5 and S7). Our exogenous peptide was used to separately compete with two peptides expressed on the surface of yeast. We expressed $CLIP_{81\text{-}101}$ peptide (LPKPPKPVSKMRMATP LLMQA) on yeast, as well as a synthetic HLA-DP401 binder, which was the most frequent peptide in the inferred central binding register from our HLA-DP401 selections (KFFLLLPMCVWCK). The synthetic strong binder matches with the overall motif of enriched peptides, with hydrophobic anchor residues. We assessed binding of HLA-DP401 to each of the yeast-expressed peptides in the presence of varying concentrations of exogenously added $CLIP_{81\text{-}101}$ peptide. We see a clear loss of MHC tetramer signal as we increase the concentration of competitor peptide, with the synthetic strong

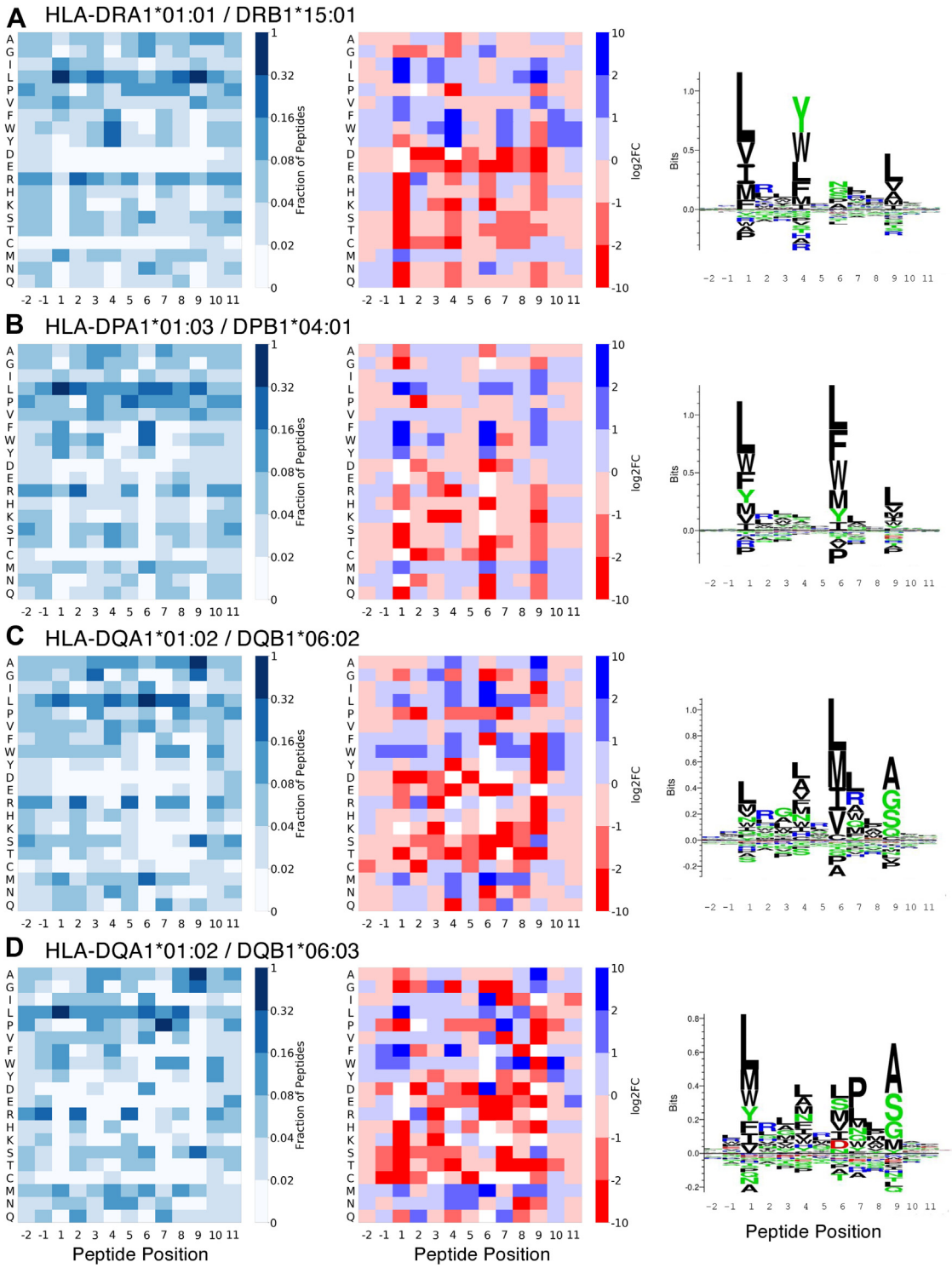

Figure 4. Motifs for peptides enriched for additional MHC-II alleles. Motifs of enriched peptides in the central register for *A* HLA-DRB1*15:01 (HLA-DR15), *B* HLA-DPA1*01:03/DPB1*04:01 (HLA-DP401), *C* HLA-DQA1*01:02/DQB1*06:02 (HLA-DQ0602), and *D* HLA-DQA1*01:02/DQB1*06:03 (HLA-DQ0603). *Left*: Heatmap highlighting the fraction of peptides with each amino acid at each position. *Center*: Log_2 fold change of positional amino acid frequency compared with unselected library. Amino acids that did not appear in the enriched sequences at a given position are *white*. *Right*: Sequence logo of enriched peptides.

binder retaining signal at higher concentrations of competitor than the yeast-expressed CLIP peptide. Addition of exogenous peptide, such as a weak and pan-MHC-II allele binder like CLIP (24), presents an opportunity to further segregate weak and strong binders in selections and another method for tuning the stringency of selection experiments.

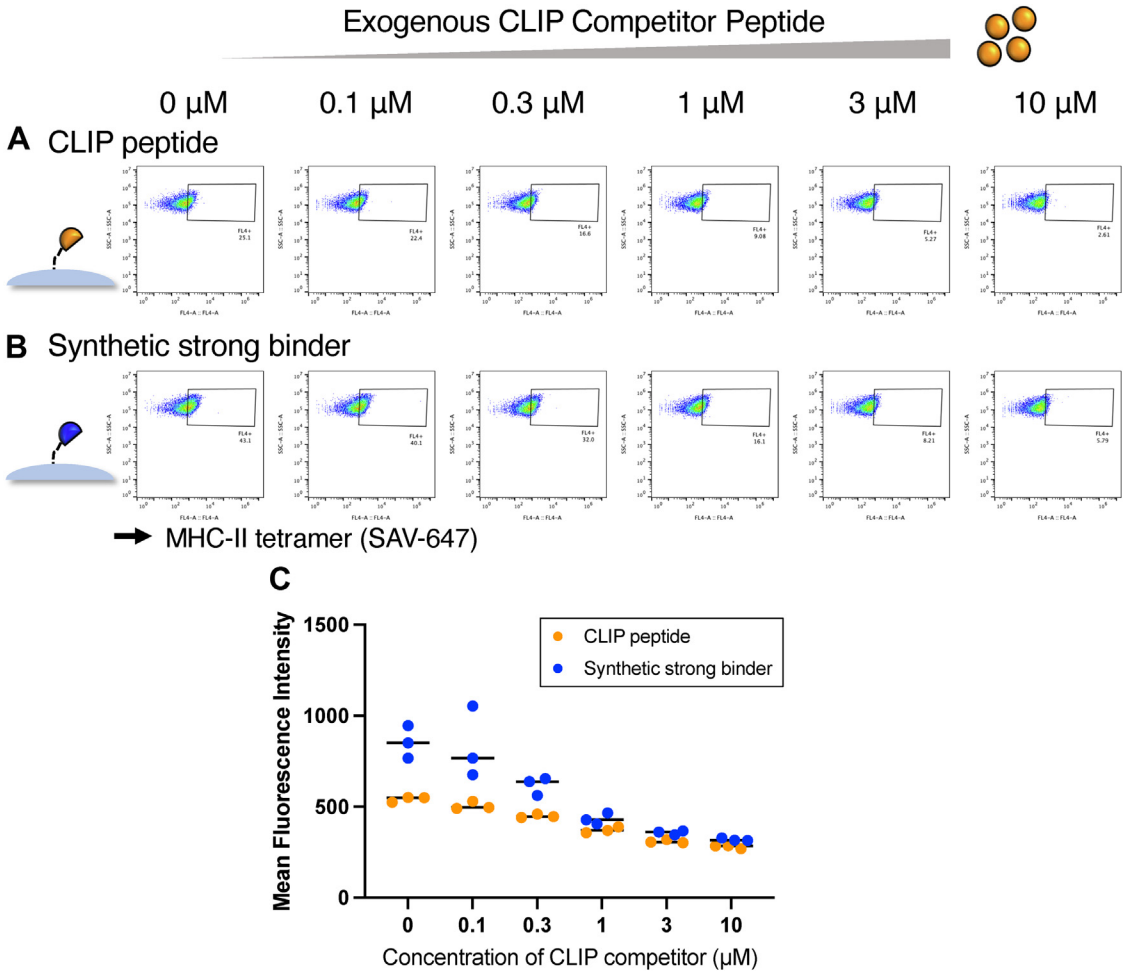

Figure 5. Competition with exogenous competitor peptide. Binding of tetramerized HLA-DP401 to *A* $CLIP_{81\text{-}101}$ peptide or *B* synthetic strong binding peptide expressed on yeast, and with exogenous $CLIP_{81\text{-}101}$ peptide added as a competitor, from 0 μM to 10 μM. Data from replicate experiments are summarized in *C*.

Discussion

Yeast-displayed peptide-MHC constructs have proven useful for assessing peptide-MHC and pMHC-TCR binding but have been limited by the need to individually validate and optimize MHCs for yeast expression (1, 4, 8). To overcome this challenge, we have developed a yeast surface display platform that decouples peptide and MHC expression, enabling users to express MHCs without optimizing them for yeast expression. Libraries of peptides are expressed on the surface of yeast, and binding to recombinantly expressed MHCs is assessed. Previous yeast display approaches decoupling peptide and MHC expression have expressed both peptide and MHC in yeast (9), which still presents the limitations of yeast protein folding, whereas expressing MHC-II proteins in an orthogonal expression system such as insect cells enables more reliable MHC-II protein folding. Separating the MHC from yeast expression also enables users to make a single library that can be screened against any recombinantly expressed MHC-II protein, so assessing a new MHC does not require generation of a new yeast display library. Here we demonstrated the principle of the platform (Fig. 1) and have shown utility for generating datasets across HLA-DR, HLA-DP, and HLA-DQ alleles (Figs. 2 and 4). Using a single peptide library and expressing these MHCs recombinantly, we were able to quickly generate high-quality, large-scale datasets of peptide binders.

Our results, which demonstrate linear peptides displayed on the surface of yeast can be used to identify MHC-binding motifs, compare well with previous studies utilizing phage display to investigate peptide-MHC binding (25–33). In these works, phage-displayed peptides were assessed for binding to MHC protein, and the identities of MHC-binding peptides were determined *via* Sanger sequencing of individual enriched clones (25, 26). While there are similarities in the implementations and results of these approaches, there are also notable differences. First, the use of next-generation sequencing in our approach enables thorough examination and identification of thousands of peptides and the determination of MHC-binding motifs at a larger scale. While phage display-generated motifs were able to capture preferences at a subset of peptide anchors (25, 27, 29, 30), the yeast display-generated motifs described here more closely resemble datasets generated through orthogonal high-throughput methods, such as mass spectrometry (14–16). Differences between the

surface density of displayed peptides on phage and yeast may also contribute to differences in identified motifs (34). The yeast display–based approach also enables determination of MHC binding *via* fluorescently labeled MHC molecules analyzed by flow cytometry. This both provides a straightforward means for examining the binding of individual peptide sequences and can be used to select and identify high- *versus* low-avidity MHC-binding sequences in a library setting *via* fluorescence-activated cell sorting (FACS).

Peptide-MHC-II binding in the peptide display platform captures a composite of binding dynamics. Namely, peptide display requires dissociation of the invariant CLIP peptide from the MHC groove and binding of other peptides. As a result, we can capture the interplay of CLIP dissociation and peptide binding to MHC, and potential equilibrium binding and unbinding. In contrast, many previous yeast display library profiling systems capture only peptide dissociation since peptides of interest begin bound to the MHC-II (3, 4, 7). In addition, the clear dependence on HLA-DM presence (Fig. S2*B*) stands in contrast to the minimal HLA-DM reliance in a dissociation-based, yeast display pMHC-II platform, which has previously shown similar motifs in library selections with or without HLA-DM (4), suggesting HLA-DM may be required for dissociation of CLIP and association of test peptides. This is consistent with phage display experiments in which HLA-DM was observed to facilitate dissociation of phage-displayed CLIP peptide from HLA-DR binding (28).

Selection strategies using peptide display are versatile and tunable. For instance, we demonstrate the use of tetramerized MHC-II proteins (Fig. 2) as well as monomeric protein for selections (Fig. S5) and the use of exogenous competitor peptide for tuning binding stringency (Fig. 5). Furthermore, there is great flexibility for use of MHC molecules. Essentially any source of recombinantly expressed class II MHCs with exchangeable peptides can be utilized, including commercial sources used for tetramers (35). Production should be agnostic to use of bacterial (36), insect (1), or other mammalian expression systems (37, 38). We also explored constraints of the system, including reliance on 3C protease, HLA-DM, and acidic conditions (Figs. S2 and S3). Further exploration of the construct itself can help tune the platform for downstream applications and ensure minimal construct-specific artifacts.

Here, we utilized CLIP peptides to stabilize recombinant MHC and mimic the stabilization of MHC seen during native antigen processing (39). However, the identity of the stabilizing peptide can be altered to change its affinity for MHC, which may tune the stringency of peptide-MHC selections. Stabilizing peptide likely sets an affinity threshold for binding, which yeast-displayed peptides must surpass. Furthermore, inclusion of stabilizing peptide likely increases the uniformity and reproducibility of selections by preventing MHC from randomly binding a distribution of peptides during expression.

The peptide display platform may impact multiple application areas. Data from high-throughput yeast display screens have shown benefit for training prediction algorithms (4), and the generalizability of the peptide display platform across MHC alleles can help expand data generation potential. In addition, given the extensibility across alleles and possibility to combine with defined peptide libraries (6, 7), this platform would be ideal for generating and assessing binding of whole human or bacterial proteomes to MHC, and potentially pMHC complexes to TCR, for applications in antigen discovery in infectious disease, autoimmunity, cancer, and transplantation.

Experimental procedures

Vector formatting for peptide display

pCT302 vector (10) was adapted to fuse a peptide sequence to Aga2. We left unchanged the N terminus of the construct, in which Aga2 is connected *via* a linker to an HA epitope tag (YPYDVPDYA) (10). We also left unchanged the subsequent C-terminal linker sequence (LQASGGGGSGGGGSGGGGSAS) (10). The C terminus of this linker is connected to the N terminus of our peptide sequence of interest. To assess frameshift mutations in our peptide sequence, we link the peptide to a Myc epitope tag (EQKLISEEDL) *via* a short linker (GGSGG). The Myc tag is proceeded by a stop codon. pCT302 was a gift from Dane Wittrup.

Peptide display parameter exploration experiments

To optimize peptide display reaction conditions, we generated reaction mixtures containing MHC-II tetramers, HLA-DM, and 3C protease in PBS. Yeast were washed with PBS then resuspended in the reaction mixture made in PBS. Yeast were incubated for 45 min at room temperature, rotating, then washed into FACS buffer (0.5% bovine serum albumin and 2 mM EDTA in 1× PBS) and assessed *via* flow cytometry (Accuri C6 flow cytometer, BD Biosciences). In the reaction mixture for MHC-II titration, we included 1 μM 3C protease and 3.5 μM of HLA-DM. In the reaction mixture for HLA-DM titration, we included 1 μM 3C protease and 75 nM MHC tetramer. And in the reaction mixture for 3C protease titration, we included 0.5 μM HLA-DM and 75 nM MHC tetramer.

For testing individual peptides with measured affinity values (Fig. 3), similar conditions were utilized: 75 nM HLA-DR401 tetramer, 1 μM 3C, and 0.5 μM HLA-DM. Yeast were also subsequently incubated in an excess of acidic saline and sampled after 45 min. Experiments assessing the impacts of CLIP as a competitor peptide (Fig. 5) utilized the same concentrations, with 75 nM HLA-DP401 tetramer, 1 μM 3C, 0.5 μM HLA-DM, in addition to 10 μM TCEP and 20-min preincubation of yeast with TCEP-containing PBS before incubating with the reaction mixture. Yeast were sampled after a 45-min room temperature incubation. Gating was applied based on forward and side scatter (SSC-A and FSC-A) on unstained yeast. Mean fluorescence intensity values were calculated using FlowJo 10.8.1 and plotted using GraphPad Prism 9.4.1.

MHC tetramers were generated using streptavidin–Alexa Fluor 647 (SAV-647; made in-house as described (40)), with a 5-fold molar ratio of MHC to SAV-647, which has four binding sites. Tetramer concentration is given as the concentration of SAV-647, representing the concentration of tetrameric species.

For gel-based assessment of cleavage of the CLIP-HLA-DRβ linker (Fig. S1), 3C protease and HLA-DR401 were coincubated at a ratio matching library selection for 45 min at room temperature before being loaded on a 12% polyacrylamide gel and run for 45 min at 180 V. The gel was stained with Coomassie stain (0.05% weight/volume Brilliant Blue R-250, 50% ethanol, 10% glacial acetic acid) and destained (50% ethanol, 10% glacial acetic acid). Mark12 protein ladder (Thermo Fisher) was included for comparison.

Randomized library generation

A 13mer randomized peptide library was generated by performing polymerase chain reaction (PCR) with a degenerate primer encoding NNK codons (N = A, C, T, or G; K = G or T). We utilized a stop codon–containing template for PCRs to decrease the likelihood of contamination with a strong-binding construct. Libraries were generated by electroporating RJY100 yeast (41) with a 5:1 mass ratio of randomized peptide product and linearized pCT302 vector.

Recombinant protein expression

Proteins were expressed recombinantly, as previously (4, 7). The extracellular domains of the α and β chains from HLA-DR401, HLA-DR15, HLA-DQ0602, HLA-DQ0603, HLA-DP401, and HLA-DM were cloned into pAcGP67a vectors for expression in insect cells. The ectodomains of each chain were formatted with a poly-histidine purification site and encoded on separate plasmids. The α chains contained a C-terminal AviTag biotinylation site (GLNDIFEAQKIEWHE), for site-specific biotinylation. Excepting HLA-DM, the N termini of β chains were linked to $CLIP_{87\text{-}101}$ peptide (for HLA-DR alleles) or an extended $CLIP_{81\text{-}101}$ peptide (for HLA-DP and HLA-DQ alleles) *via* a flexible Gly-Ser linker containing a 3C protease site (LEVLFQGP). Plasmid DNA was transfected into SF9 insect cells with BestBac 2.0 DNA (Expression Systems) using Cellfectin II reagent (Thermo Fisher). Viruses for α and β chains were cotitrated to determine an optimal ratio for heterodimer formation, cotransduced into High Five (Hi5) insect cells (Thermo Fisher), and incubated for 48 to 72 h. Proteins were secreted by Hi5s and purified by conditioning the cell culture media with 50 mM Tris-HCl (pH 8), 1 mM $NiCl_2$, and 5 mM $CaCl_2$, clearing the precipitant *via* centrifugation, and incubating the resulting supernatant with Ni-NTA resin (1). Proteins were further purified using an S200 increase column on an AKTAPURE FPLC (GE Healthcare). Excepting HLA-DM, protein was site specifically biotinylated and confirmed using a gel-shift assay (42).

Library selections

The randomized 13mer library had 1×10^8 unique transformants, and a random subsampling of 5×10^6 yeast, based on absorbance, was grown up for use in selections in order to increase coverage of enriched yeast by sequencing and for easier comparison across iterations of selections. An overrepresentation of these yeast were input into round 1, with 5×10^7 yeast as input, measured by absorbance. Rounds 2 and 3 used 2.5×10^7–5×10^7 input yeast.

To perform tetrameric selections, yeast were first washed into pH 7.2 PBS with 10 μM TCEP and incubated for 20 min while making the reaction mixture. The reaction mixture contained MHC tetramers, generated by preincubating 75 nM SAV-647 and 375 nM biotinylated MHC. Tetramers are added to the appropriate volume of pH 7.2 PBS with 10 μM TCEP. This is followed by the addition of 3C protease to 1 μM, then the addition of 0.5 μM HLA-DM (excepting round 3 of HLA-DR401 no-streptavidin selections, which received 1 μM HLA-DM). TCEP was added to prevent disulfide bonds, which could diminish the enrichment of peptides containing cysteine. Yeast were spun down and resuspended in reaction mixture at 2×10^8 yeast/ml. This reaction was incubated at room temperature for 45 min, rotating, covered to decrease risk of photobleaching. Then, yeast in the reaction mixture were placed on ice and a subsample was taken for assessment *via* flow cytometry, and the remaining yeast were moved to a 9-fold excess of pH 5 citric acid saline buffer (20 mM citric acid, 150 mM NaCl). Yeast were incubated in the acid saline at 4 °C for 45 min, rotating. Yeast were then washed into FACS buffer, again at 2×10^8 yeast/ml, and ∝-Alexa Fluor 647 beads (Miltenyi Biotec) equal to 10% of the FACS buffer volume were added. Yeast were incubated with beads at 4 °C, rotating for 15 min. Yeast were then pelleted, resuspended in 5 ml FACS buffer, run over a Miltenyi LS column (Miltenyi Biotec), washed three times with 3 ml FACS buffer, and eluted with 5 ml FACS buffer. The elution contained the yeast that were bound to the MHC. Selected yeast were then washed into SDCAA medium and grown to confluence. Yeast were subcultured in SGCAA medium at $A_{600} = 1$ for 48 to 72 h at 20 °C before each experiment (43).

Monomeric selections were performed similarly to tetrameric selections, with a few modifications. Monomeric MHC-II was added in the reaction mixture, absent SAV-647. MHC-II, 300 nM, was added, equivalent to the amount of MHC-II that binds to the four binding sites of 75 nM SAV-647 in tetrameric selections. Streptavidin beads (Miltenyi Biotec) were added in place of ∝-Alexa Fluor 647 beads, and selections proceeded in the same manner.

Deep sequencing and data processing

Peptide identities were determined from enriched yeast following selections. Plasmid DNA was extracted from 5×10^7 input yeast *via* the Zymoprep Yeast Miniprep Kit (Zymo Research) following manufacturer's instructions. Amplicons were generated by two rounds of PCR to add inline i5 and i7 paired-end handles and inline sequencing barcodes, unique to rounds multiplexed in a single sample. PCR primers were designed to capture the 13mer peptide sequence, priming off of adjacent constant sequences. Amplicons were sequenced with an Illumina MiSeq (Illumina Incorporated) at the MIT BioMicroCenter using 300 nucleotide kits.

Paired-end reads were assembled using PandaSeq (44). Given the large space of possible NNK-encoded 13mer peptides, single-nucleotide variants are likely to be a result of PCR or sequencing errors. Using CDHIT (45), single-nucleotide

variants were combined and the more frequent sequence was chosen as the consensus sequence.

Peptide binding register inference and motif visualization

To infer the register in which the enriched 13mer peptides bound, we utilized a position weight matrix (PWM) method (7). Enriched peptides from round 3 of selection were one-hot encoded and padded on the C terminus of the peptide. We utilized a 9mer PWM and eight registers and one "trash" register.

At the start, peptides were randomly assigned to registers and a 9mer PWM was generated. The contribution of a given peptide to the PWM was weighted by its read count. Over successive iterations, peptides were assigned to new registers and the PWM was updated, excluding peptides in the trash cluster. Assignment to each register is stochastic, but biased, such that registers that match the PWM were favored. At each assignment, if the peptide was assigned to a nontrash register, we first took out that peptide from the PWM. The PWM then defined an energy value for each register option, which was used to generate a Boltzmann distribution from which the updated register shift was sampled. Stochasticity was decreased over time by raising the inverse temperature linearly from 0.05 to 1 over 20 iterations, simulating cooling (46). The final iteration was deterministic, where the distribution concentrated solely on the locally optimal register option.

From selections with HLA-DR401 (tetramers), HLA-DR401 (monomers), HLA-DP401, HLA-DR15, HLA-DQ0602, and HLA-DQ0603, we identified 11,275/32,885/89,746/46,276/27,231/8289 enriched peptides in our sequencing after round 3, respectively. Of these, from register inference, we identified 6164/18,845/56,977/28,167/11,584/2219 peptides in native registers, respectively.

Heatmap and sequence logo visualizations were generated for peptides that were inferred to bind in the central register, with two peptide flanking residues on either side of the 9mer core. Heatmaps were generated using a custom script to show amino acid frequency or $\log_2$ fold change over the unselected library at each position. Amino acids that did not appear at a given position in the round 3 library appear as white spaces in the log2 fold change heatmap. Sequence logos were generated using Seq2Logo (47), with default settings, except background frequencies were generated from the unselected library. In heatmaps and sequence logos visualizations, peptides were not weighted by read counts.

Data availability

All deep sequencing files are deposited on the Sequence Read Archive (SRA) with accession code PRJNA867335 (http://www.ncbi.nlm.nih.gov/bioproject/867335). Round 3 enriched peptides with stop codon–containing peptides removed are available in Supporting Information 1. Inferred register clusters for enriched peptides are indicated, with the trash register indicated as register "-1." Peptides from the unselected library are also provided in Supporting Information 1.

Code availability

Scripts used for data processing, register inference, and visualization are publicly available at https://github.com/birnbaumlab/Huisman-et-al-2022-Peptide-Display.

Supporting information—This article contains supporting information.

Acknowledgments—We would like to thank Zheng Dai and David Gifford for helpful discussions and thoughtful review of the manuscript. We thank the MIT BioMicro Center for performing Illumina library sequencing. This work was supported in part by the Koch Institute Support (core) Grant P30-CA14051 from the NCI, National Institutes of Health. This work was supported by the National Institutes of Health (U19-AI110495), the Melanoma Research Alliance, and the Packard Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions—B. D. H., M. E. B. conceptualization; B. D. H. methodology; B. D. H. validation; B. D. H., P. A. B. formal analysis; B. D. H., P. A. B. investigation; B. D. H., P. A. B. data curation; B. D. H., M. E. B. writing – original draft; B. D. H., P. A. B., M. E. B. writing – review & editing; B.D.H. visualization; M. E. B. supervision; M. E. B. funding acquisition.

Funding and additional information—This work was supported by a Schmidt Futures grant to M. E. B. and a National Science Foundation Graduate Research Fellowship to B. D. H.

Conflict of interest—M. E. B. is an equity holder in 3T Biosciences and is a cofounder, equity holder, and consultant of Kelonia Therapeutics and Abata Therapeutics.

Abbreviations—The abbreviations used are: MHC, major histocompatibility complex; PWM, position weight matrix; SAV, streptavidin; TCR, T-cell receptor.

References

1. Birnbaum, M. E., Mendoza, J. L., Sethi, D. K., Dong, S., Glanville, J., Dobbins, J., *et al.* (2014) Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087
2. Gee, M. H., Han, A., Lofgren, S. M., Beausang, J. F., Mendoza, J. L., Birnbaum, M. E., *et al.* (2018) Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell* **172**, 549–563.e16
3. Wen, F., Esteban, O., and Zhao, H. (2008) Rapid identification of CD4+ T-cell epitopes using yeast displaying pathogen-derived peptide library. *J. Immunol. Met.* **336**, 37–44
4. Rappazzo, C. G., Huisman, B. D., and Birnbaum, M. E. (2020) Repertoire-scale determination of class II MHC peptide binding *via* yeast display improves antigen prediction. *Nat. Commun.* **11**, 4414
5. Liu, R., Jiang, W., and Mellins, E. D. (2021) Yeast display of MHC-II enables rapid identification of peptide ligands from protein antigens (RIPPA). *Cell. Mol. Immunol.* **18**, 1847–1860
6. Dai, Z., Huisman, B. D., Zeng, H., Carter, B., Jain, S., Birnbaum, M. E., *et al.* (2021) Machine learning optimization of peptides for presentation by class II MHCs. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btab131
7. Huisman, B. D., Dai, Z., Gifford, D. K., and Birnbaum, M. E. (2022) A high-throughput yeast display approach to profile pathogen proteomes for MHC-II binding. *Elife*. https://doi.org/10.7554/eLife.78589

8. Fernandes, R. A., Li, C., Wang, G., Yang, X., Savvides, C. S., Glassman, C. R., *et al.* (2020) Discovery of surrogate agonists for visceral fat Treg cells that modulate metabolic indices *in vivo*. *Elife*. https://doi.org/10.7554/eLife.58463
9. Jiang, W., and Boder, E. T. (2010) High-throughput engineering and analysis of peptide binding to class II MHC. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13258–13263
10. Midelfort, K. S., Hernandez, H. H., Lippow, S. M., Tidor, B., Drennan, C. L., and Wittrup, K. D. (2004) Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J. Mol. Biol.* **343**, 685–701
11. Busch, R., Rinderknecht, C. H., Roh, S., Lee, A. W., Harding, J. J., Burster, T., *et al.* (2005) Achieving stability through editing and chaperoning: regulation of MHC class II peptide binding and expression. *Immunol. Rev.* **207**, 242–260
12. Jiang, W., Strohman, M. J., Somasundaram, S., Ayyangar, S., Hou, T., Wang, N., *et al.* (2015) pH-susceptibility of HLA-DO tunes DO/DM ratios to regulate HLA-DM catalytic activity. *Sci. Rep.* **5**, 17333
13. Hennecke, J., and Wiley, D. C. (2002) Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J. Exp. Med.* **195**, 571–581
14. Abelin, J. G., Harjanto, D., Malloy, M., Suri, P., Colson, T., Goulding, S. P., *et al.* (2019) Defining HLA-II ligand processing and binding rules with mass spectrometry enhances cancer Epitope prediction. *Immunity* **51**, 766–779.e17
15. Scally, S. W., Petersen, J., Law, S. C., Dudek, N. L., Nel, H. J., Loh, K. L., *et al.* (2013) A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J. Exp. Med.* **210**, 2569–2582
16. Racle, J., Michaux, J., Rockinger, G. A., Arnaud, M., Bobisse, S., Chong, C., *et al.* (2019) Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* **37**, 1283–1286
17. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454
18. Neefjes, J., Jongsma, M. L. M., Paul, P., and Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836
19. Pugliese, A., Boulware, D., Yu, L., Babu, S., Steck, A. K., Becker, D., *et al.* (2016) HLA-DRB1*15:01-DQA1*01:02-DQB1*06:02 haplotype protects autoantibody-positive relatives from type 1 diabetes throughout the stages of disease progression. *Diabetes* **65**, 1109–1119
20. Kaushansky, N., and Ben-Nun, A. (2014) DQB1*06:02-associated pathogenic anti-myelin autoimmunity in multiple sclerosis-like disease: potential function of DQB1*06:02 as a disease-predisposing allele. *Front. Oncol.* **4**, 280
21. Karnes, J. H., Bastarache, L., Shaffer, C. M., Gaudieri, S., Xu, Y., Glazer, A. M., *et al.* (2017) Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.aai8708
22. Robinson, J., Georgiou, X., Flicek, P., and The, I.-I. (2020) SGE the IPD-IMGT/HLA database nucleic acids Research. *Database Nucl. Acids Res.* **43**, D948–D955
23. Jiang, W., Birtley, J. R., Hung, S.-C., Wang, W., Chiou, S.-H., Macaubas, C., *et al.* (2019) *In vivo* clonal expansion and phenotypes of hypocretin-specific CD4+ T cells in narcolepsy patients and controls. *Nat. Commun.* **10**, 5247
24. Fallang, L.-E., Roh, S., Holm, A., Bergseng, E., Yoon, T., Fleckenstein, B., *et al.* (2009) Complexes of two cohorts of CLIP peptides and HLA-DQ2 of the autoimmune DR3-DQ2 haplotype are poor substrates for HLA-DM. *J. Immunol.* **182**, 726.1–72726
25. Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B., *et al.* (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. *Cell* **74**, 197–203
26. Hammer, J., Takacs, B., and Sinigaglia, F. (1992) Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J. Exp. Med.* **176**, 1007–1013
27. Hammer, J., Belunis, C., Bolin, D., Papadopoulos, J., Walsky, R., Higelin, J., *et al.* (1994) High-affinity binding of short peptides to major histocompatibility complex class II molecules by anchor combinations. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4456–4460
28. Raddrizzani, L., Bono, E., Vogt, A. B., Kropshofer, H., Gallazzi, F., Sturniolo, T., *et al.* (1999) Identification of destabilizing residues in HLA class II-selected bacteriophage display libraries edited by HLA-DM. *Eur. J. Immunol.* **29**, 660–668
29. Davenport, M. P., Quinn, C. L., Valsasnini, P., Sinigaglia, F., Hill, A. V., and Bell, J. I. (1996) Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. *Immunology* **88**, 482–486
30. Gregori, S., Bono, E., Gallazzi, F., Hammer, J., Harrison, L. C., and Adorini, L. (2000) The motif for peptide binding to the insulin-dependent diabetes mellitus-associated class II MHC molecule I-Ag7 validated by phage display library. *Int. Immunol.* **12**, 493–503
31. Fujisao, S., Matsushita, S., Nishi, T., and Nishimura, Y. (1996) Identification of HLA-DR9 (DRB1*0901)-binding peptide motifs using a phage fUSE5 random peptide library. *Hum. Immunol.* **45**, 131–136
32. Stratmann, T., Apostolopoulos, V., Mallet-Designe, V., Corper, A. L., Scott, C. A., Wilson, I. A., *et al.* (2000) The I-Ag7 MHC class II molecule linked to murine diabetes is a promiscuous peptide binder. *J. Immunol.* **165**, 3214–3225
33. Milik, M., Sauer, D., Brunmark, A. P., Yuan, L., Vitiello, A., Jackson, M. R., *et al.* (1998) Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat. Biotechnol.* **16**, 753–756
34. Boder, E. T., and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557
35. Davis, M. M., Altman, J. D., and Newell, E. W. (2011) Interrogating the repertoire: broadening the scope of peptide-MHC multimer analysis. *Nat. Rev. Immunol.* **11**, 551–558
36. Altman, J. D., Reay, P. A., and Davis, M. M. (1993) Formation of functional peptide complexes of class II major histocompatibility complex proteins from subunits produced in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10330–10334
37. Willis, R. A., Ramachandiran, V., Shires, J. C., Bai, G., Jeter, K., Bell, D. L., *et al.* (2021) Production of Class II MHC proteins in Lentiviral vector-transduced HEK-293T cells for tetramer staining reagents. *Curr. Protoc.* **1**, e36
38. Day, C. L., Seth, N. P., Lucas, M., Appel, H., Gauthier, L., Lauer, G. M., *et al.* (2003) *Ex vivo* analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J. Clin. Invest.* **112**, 831–842
39. Roche, P. A., and Furuta, K. (2015) The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* **15**, 203–216
40. Ramachandiran, V., Grigoriev, V., Lan, L., Ravkov, E., Mertens, S. A., and Altman, J. D. (2007) A robust method for production of MHC tetramers with small molecule fluorophores. *J. Immunol. Met.* **319**, 13–20
41. Van Deventer, J. A., Kelly, R. L., Rajan, S., Wittrup, K. D., and Sidhu, S. S. (2015) A switchable yeast display/secretion system. *Protein Eng. Des. Sel.* **28**, 317–325
42. Fairhead, M., and Howarth, M. (2015) Site-specific biotinylation of purified proteins using BirA. *Met. Mol. Biol.* **1266**, 171–184
43. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768
44. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform.* **13**, 31
45. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152
46. Andreatta, M., Alvarez, B., and Nielsen, M. (2017) GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucl. Acids Res.* **45**, W458–W463
47. Thomsen, M. C. F., and Nielsen, M. (2012) Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–W287