High-throughput identification of MHC class I binding peptides using an ultradense peptide array*

Amelia K. Haj[1], Meghan E. Breitbach[1], David A. Baker[1], Mariel S. Mohns[1], Gage K. Moreno[1], Nancy A. Wilson[2], Victor Lyamichev[3], Jigar Patel[3], Kim L. Weisgrau[4], Dawn M. Dudley[1], David H. O'Connor[1]

[1]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison

[2]Department of Medicine, University of Wisconsin-Madison

[3]Nimble Therapeutics, Madison, WI

[4]Wisconsin National Primate Research Center, Madison, WI

* This work was supported by R24OD017850 and National Library of Medicine Grant 5T15LM007359.

Abstract

Rational vaccine development and evaluation requires identifying and measuring the magnitude of epitope-specific CD8 T cell responses. However, conventional CD8 T cell epitope discovery methods are labor-intensive and do not scale well. Here, we accelerate this process by using an ultradense peptide array as a high-throughput tool for screening peptides to identify putative novel epitopes. In a single experiment, we directly assess the binding of four common Indian rhesus macaque MHC class I molecules – Mamu-A1*001, -A1*002, -B*008, and -B*017 – to approximately 61,000 8-mer, 9-mer, and 10-mer peptides derived from the full proteomes of 82 simian immunodeficiency virus (SIV) and simian-human immunodeficiency virus (SHIV) isolates. Many epitope-specific CD8 T cell responses restricted by these four MHC molecules have already been identified in SIVmac239, providing an ideal dataset for validating the array; up to 64% of these known epitopes are found in the top 192 SIVmac239 peptides with the most intense MHC binding signals in our experiment. To assess whether the peptide array identified putative novel CD8 T cell epitopes, we validated the method by IFN-γ ELISPOT assay and found three novel peptides that induced CD8 T cell responses in at least two Mamu-A1*001-positive animals; two of these were validated by *ex vivo* tetramer staining. This high-throughput identification of peptides that bind class I MHC will enable more efficient CD8 T cell response profiling for vaccine development, particularly for pathogens with complex proteomes where few epitope-specific responses have been defined.

Introduction

CD8 T cell receptors typically recognize epitopes – linear peptides of 8-11 amino acids – presented by major histocompatibility complex (MHC) class I molecules. These peptides are usually derived from self proteins that under normal circumstances do not induce an immune response. Cancerous cells, transplanted cells, and infected cells, however, present non-self peptides that are recognized as foreign by CD8 T cells (1). CD8 T cell recognition of foreign epitopes, including those derived from viral proteins, along with concurrent recognition of costimulatory molecules on the antigen-presenting cell, leads to T cell activation and production of cytotoxic compounds that trigger apoptosis in the affected cell (2). Polymorphisms in the peptide binding cleft of the MHC glycoprotein determine the variety of peptides each molecule is able to present, and specificity is determined in part by anchor residues in the peptide binding to specific pockets in the binding cleft (3). Identifying which peptides are presented by particular MHC molecules and are likely to elicit epitope-specific CD8 T cell responses is a critical task in designing prophylactic and therapeutic vaccines against infectious diseases and cancers (4, 5).

Characterizing MHC-peptide binding can be a time- and resource-intensive task. Competitive binding assays, which calculate the concentration (IC_{50}) at which a peptide of interest outcompetes 50% of a radiolabeled probe peptide for binding to MHC class I molecules, are used to estimate MHC-peptide binding affinity (6). These methods are low-throughput and difficult to scale. Computational tools have been developed that have the potential to improve throughput, enabling prediction of the likelihood a given peptide will bind to a specific MHC molecule. However, these methods exploit our understanding of anchor residues and rely on previously generated peptide binding datasets for pattern recognition, and so are limited to MHC molecules whose binding repertoires have already been well studied, creating a self-fulfilling prophecy of sorts (6, 7). A different computational technique, molecular docking, uses the three-dimensional

61 crystal structure of MHC molecules to predict peptides likely to bind (8, 9). While this method has
62 the benefit of working for both MHC class I and II molecules, it still cannot serve as a substitute
63 for *in vitro* binding assays.
64
65 CD8 T cell epitope identification is a similarly involved task, often requiring large quantities of cells
66 exposed *in vivo* to the pathogen of interest. In our previous work identifying CD8 T cell epitopes,
67 we performed IFN-γ ELISPOTs using pools of 15-mer peptides to stimulate peripheral blood
68 mononuclear cells (PBMCs), followed by successive rounds of deconvolution of positive wells to
69 determine stimulatory peptides and minimal optimal epitopes within positive 15-mers (10). This is
70 followed by generation of MHC-peptide tetramers, which stably bind T cell receptors and can be
71 used for cell staining and sorting (11). While successful, these methods are not ideal – using 15-
72 mer peptides, which are too large to bind class I MHC *in vivo*, can lead to false negatives in
73 ELISPOT assays (12). Currently, there is no high-throughput *in vitro* tool for assessing whether
74 shorter peptides are likely to bind to MHC and therefore are more likely to be CD8 T cell epitopes.
75
76 Rhesus macaques infected with simian immunodeficiency virus (SIV) and simian-human
77 immunodeficiency virus (SHIV) have long been used as models for HIV infection studies. Rhesus
78 macaque class I MHC class I A (Mamu-A) and class I B (Mamu-B) genes bind and present
79 peptides to CD8 T cells like their HLA-A and HLA-B counterparts in humans. Previous studies
80 have characterized the peptide binding specificity of common Mamu-A and Mamu-B molecules
81 to SIV-derived epitopes, and have identified epitopes that elicit CD8 T cell responses (13–16).
82 Knowledge of these epitopes enables tracking of SIV-specific responses after infection and
83 vaccination (17), and identification of the peptide binding specificities of MHC alleles commonly
84 found in elite controllers has guided the development of epitope-based vaccines against SIV (18).
85 Yet, even in this well-characterized system, T cell epitopes have only been defined for a small
86 minority of the hundreds of known MHC class I alleles, while nearly no T cell epitopes have been

87 defined for other macaque viruses or bacterial pathogens. While considerably more effort has
88 been dedicated to defining T cell epitopes in humans, these efforts are similarly concentrated on
89 a small number of pathogens (e.g., HIV, influenza). Methods that facilitate the high-throughput
90 identification of T cell epitopes, particularly for pathogens with large proteomes where T cells play
91 an essential role in protective immunity (e.g., *Mycobacterium tuberculosis*), are urgently needed.
92
93 Ultradense peptide arrays have previously been used as a high-throughput tool for assessing
94 antibody responses to pathogens; here, we repurpose them to identify peptides that bind specific
95 MHC class I molecules. Briefly, up to six million linear peptides are synthesized on a single array
96 chip, molecules of interest – antibodies in serum, or other molecules such as MHC proteins – are
97 allowed to bind to the peptides, and binding affinity is reported in fluorescence intensity units
98 based on the detection of fluorophore-conjugated secondary antibodies bound to the molecules.
99 Our group and others have designed arrays to simultaneously assess IgG and IgM binding to
100 linear pathogen-derived peptides (19, 20). Here, we generated a peptide array with 61,066 unique
101 8-mer, 9-mer, and 10-mer peptides derived from 82 SIV and SHIV isolates, and measured binding
102 to four rhesus macaque MHC class I molecules, Mamu-A1*001, -A1*002, -B*008, and -B*017, all
103 of which have peptide binding repertoires that have been extensively studied by conventional
104 methods. While we focus our attention here on a particular well-studied strain, SIVmac239, as a
105 way of validating the array, the large capacity of the array gave us the opportunity to generate the
106 same data for several dozen isolates, creating a sizable dataset that allows for comparisons
107 among strains. We tested the top 86 SIVmac239-derived peptides identified by the array as
108 having the highest binding affinity for Mamu-A1*001 for their ability to induce IFN-γ production in
109 peripheral blood mononuclear cells isolated from SIVmac239-infected, Mamu-A1*001-positive
110 rhesus macaques. We demonstrate that peptides that exhibit strong MHC binding on our array
111 frequently correspond to CD8 T cell epitopes that have been identified by conventional assays,
112 and that our high-binding peptides are enriched for putative novel epitopes that elicit CD8 T cell

113 responses. This suggests that the MHC-peptide array can be used to predict immunologically
114 relevant MHC-peptide interactions, offering a high-throughput alternative to traditional methods.
115
116

117 ## Materials and Methods

118 **SIV/SHIV peptidome array design**

119 Peptide sequences were generated from SIV and SHIV amino acid sequences (Supplemental
120 table I). We downloaded 82 SIV and SHIV GenBank files from NCBI and extracted amino acid
121 sequences for all open reading frames into FASTA files. The GenBank SIVmac239 nef protein
122 sequence is truncated and therefore the correct full-length sequence was translated from the full
123 GenBank sequence and added manually. The FASTA files were sent to Roche NimbleGen
124 (Madison, WI; now Nimble Therapeutics) to generate the array. Non-redundant 8-mer, 9-mer, and
125 10-mer peptide sequences were generated with 1-amino acid offsets, and 5 replicates for each
126 peptide were included on the array in a 12-plex array configuration.

127

128 **Amino acid substitution array design**

129 Substitution libraries were designed by introducing all possible single amino-acid substitutions
130 and single-amino-acid deletions for the Gag CM9 epitope (CTPYDINQM) and the Tat SL8 epitope
131 (STPESANL) using all 20 L-amino acids. The library was synthesized in five replicates. For this
132 array, 1785 peptides were synthesized (9 x 21 x 5 + 8 x 21 x 5 = 1785, where 9 and 8 are the
133 peptide lengths, 21 is the number of possible amino acids/deletion at each position, and 5 is the
134 number of replicates).

135

136 **Production of MHC molecules**

137 Inclusion bodies were prepared essentially according to the method of Hutchinson, et al (21). In
138 brief, bacterial expression plasmids containing the sequence for human beta-2 microglobulin
139 (β2M) or rhesus MHC class I Mamu-A1*001, Mamu-A1*002, Mamu-B*008, or Mamu-B*017 were
140 transformed into competent BL21DE3 *E. coli* bacteria. Single colonies were grown in 1 L TYP
141 media (16 g Yeast extract, 16 g Tryptone, 5 g NaCl, 1 g K_2HPO_4) with appropriate antibiotic

142 markers, induced with IPTG during logarithmic growth phase, and allowed to express protein
143 overnight. Bacterial pellets were incubated with lysis buffer (10 mM Tris pH 8.1, 150 mM NaCl,
144 10 mM EDTA with DNAse and $MgCl_2$). Following sonication, the inclusion bodies were purified in
145 50 mM Tris pH 8.1 10 mM EDTA 100 mM NaCl buffer with 0.5% Triton X-100, washed in
146 resuspension buffer without Triton-X and the pellet dissolved in denaturing buffer. β2M inclusion
147 bodies were dissolved in 8M Urea with 50 mM Tris pH 8.1, 100 mM NaCl, 10 mM EDTA and 10
148 mM DTT. MHC heavy chains were dissolved in 6 M guanidine, 50 mM Tris pH 8.1, 100 mM NaCl,
149 10 mM EDTA with 10 mM DTT. After a Bradford assay to determine the concentration of protein,
150 inclusions were snap frozen in LN_2 and frozen at -80°C in aliquots of 30 mg for MHC heavy chain
151 and 10 mg for β2M.
152
153 **MHC complex preparation**
154 To prepare MHC I complexes for peptide array binding, 45 µl di-sodium tetraborate/sodium
155 hydroxide pH 10 buffer (Sigma-Aldrich), 10 µl 2 mg/ml β2M, and 60 µg of MHC class I heavy
156 chain stock were mixed in 400 µl water. This mixture was concentrated with AmiconUltra 0.5 ml,
157 10K centrifugal filter (Millipore) at 12,000 g for 5 min without pre-incubation. Concentrated material
158 was mixed with 400 µl 0.05X pH10 buffer to further dilute urea and 2-ME and concentrated in the
159 same Amicon filter at 12,000 g for 7 min. The concentrated sample was transferred to a 1.5 ml
160 Eppendorf tube, centrifuged at 20,000 g for 5 min to remove precipitated material, and stored at
161 4ºC before binding to the peptide array.
162
163 **MHC-peptide binding assay**
164 Peptide synthesis was accomplished through light-directed array synthesis in a Maskless Array
165 Synthesizer (MAS) using an amino-functionalized substrate as previously reported (22). Each
166 peptide replicate were randomly positioned on the array. After synthesis and de-protection, the
167 array was incubated in 0.67 µg/ml Cy5-labeled streptavidin (SA-Cy5) (Amersham), 1% alkali-

168 soluble casein (Novagen) at room temperature for 1 hour to stain technical SA-binding features
169 used for array QC and gridding. After a quick rinse in water, the array was dried by spinning in a
170 microcentrifuge equipped with an array holder and a 12-plex incubation chamber was applied to
171 the array surface. Before loading onto the array, MHC I complexes were 1:6 diluted in 20 mM tris-
172 HCl, pH7.4, 1% BSA and 6 µl aliquots were loaded into 5 chambers of a 12-plex chamber well.
173 An additional chamber was loaded without the MHC I complexes as a control. The loading ports
174 were sealed with port seals (Grace Bio-Labs, Bend, OR) and the array was incubated at 4ºC
175 overnight. After incubation the chamber was removed, and slides were rinsed in water and stained
176 by dipping with Alexa Fluor 647 MEM-123 1:1,000 (Novus Biologicals, Centennial, CO). After
177 rinsing in water and drying as described above, the array was scanned to collect fluorescence
178 data. Cy5 fluorescence intensity of the array was measured with an MS200 scanner at resolution
179 2 µm, wavelength 635 nm. Cy5 fluorescence intensities were extracted using Image Extraction
180 Software.
181
182 **Peptide array data analysis**
183 The array data contains the fluorescence intensity, reported in arbitrary signal intensity units, for
184 each peptide sequence replicate for each MHC molecule. We took the base 2 logarithm of the
185 fluorescence intensity and used the median value for each of the unique peptide
186 sequences. Using the median reduces noise induced by the peptide position on the array and
187 reduces the overall variance. We assessed the top 24, 48, 96, and 192 peptides with the highest
188 binding intensities in each virus-MHC class I molecule combination for the presence of known
189 epitopes.
190
191 **IFN-γ ELISPOT assay**
192 ELISPOTs were performed in duplicate using frozen rhesus macaque PBMCs isolated from
193 EDTA-treated whole blood. Frozen cells were thawed slowly at room temperature and washed

194 twice in R10 media containing RPMI, 10% FBS, 1% antibiotic/antimycotic, and 1% L-glutamine
195 (Hyclone, Logan, UT). Approximately 100,000 cells in 100 µl of R10 media were layered on pre-
196 coated monkey IFN-γ ELISpot-Plus plates (Mabtech Inc., Mariemont, Ohio, USA) with a single
197 peptide per well at a final concentration of 1 µM. Peptides were synthesized by GenScript
198 (Piscataway, NJ, USA). Concanavalin A (ConA, 10 µM) was used as a positive control and cells
199 with no peptide were used as a negative control.
200
201 **ELISPOT analysis**
202 Developed ELISPOT plates were read with an AID Elispot reader using the IFN-γ setting and a
203 threshold of 100 spots per million cells. ConA wells were counted using custom parameters
204 defined for ConA (Min intensity: 10, Min size: 20, Min gradient: 1, Max intensity: 57, Max size:
205 295, Max gradient: 50, Basic algorithm settings emphasis: tiny, Algorithm: C); the rest were read
206 using custom IFN-γ parameters (Min intensity: 20, Min size: 40, Min gradient: 1, Max intensity:
207 254, Max size: 5000, Max gradient: 90, Basic algorithm settings emphasis: small, Algorithm: C).
208 Positive ELISPOT responses were determined as previously described (23).
209
210 **Tetramers and tetramer staining**
211 Tetramers conjugated to APC or PE were generously produced by the NIH Tetramer Core Facility
212 using peptides generated by GenScript. Thawed PBMC resuspended to 1 x 10^6 cells/100 µl of
213 R10 medium were stained with tetramers at a concentration of 5 µg/ml in the dark for one hour at
214 37ºC, and were subsequently stained with CD3 Alexa Fluor 700 (BD Biosciences, San Jose, CA)
215 and CD8 BV421 (BD Biosciences) for 15 minutes at room temperature, washed, and fixed with
216 2% paraformaldehyde. Samples were run on a BD LSRII flow cytometer within 24 hours and were
217 analyzed with FlowJo software, version 10.5.3.
218
219

220 # Results

221 **Identification of SIVmac239 peptides that bind MHC class I molecules**

222 We tested binding of purified rhesus MHC class I Mamu-A1*001, -A1*002, -B*008, and -B*017
223 molecules to the full proteomes of 82 SIV and SHIV isolates (Supplemental table I). Previous
224 studies have identified SIVmac239 epitopes that bind each of the four MHC class I molecules
225 tested here, which enables us to validate the peptide array as a method for screening for putative
226 novel CD8 T cell epitopes (14–16, 24). To determine whether the array identified previously-
227 established SIVmac239 epitopes, we determined the percentage of known CD8 T cell epitopes
228 that were identified in the top 24, 48, 96, and 192 (convenient fractions/multiples of a 96-well
229 ELISPOT plate) highest-binding peptides for each MHC molecule. The top 192 peptides for each
230 of the four MHC molecules encompassed 27-64% of the known T cell epitopes restricted by each
231 molecule (Table I). Figure 1 shows the complete data for SIVmac239 peptides binding to Mamu-
232 A1*001, -A1*002, -B*008, and -B*017, with previously-identified epitopes present in the top 192
233 peptides (out of 9912 SIVmac239 peptides, or 1.9%) for each molecule indicated.

234

235 **Critical residues for Mamu-A1*001 peptide binding identified through substitution analysis**

236 Peptide anchor residues define the binding motif for a given MHC molecule and enable prediction
237 of other peptides likely to bind. The binding motif for Mamu-A1*001 was previously defined by
238 sequence analysis of peptides bound to the molecule in vitro and by performing competitive
239 binding assays using peptides with amino acid substitutions (24, 25). These studies identified
240 positions two and three (P2 and P3), and the C-terminus, as important for binding. We performed
241 a full substitution analysis at each amino acid for two known immunodominant CD8 T cell epitopes
242 that bind to Mamu-A1*001, Gag CM9 (Gag 181-189) and Tat SL8 (Tat 28-35). We tested the
243 influence of all 20 amino acids, plus an amino acid deletion, at each position on the binding of the
244 peptides to Mamu-A1*001 molecules. As would be expected based on prior studies, substitutions

245 at P2 (threonine) and P3 (proline) for both peptides were generally poorly tolerated and resulted
246 in dramatically decreased signal intensity, indicating reduced binding affinity (Figure 2). At P2, a
247 substitution with serine was tolerated; P3 did not tolerate any substitutions. These results
248 corroborated those found by competitive binding assays using Gag CM9 peptides with single
249 amino acid substitutions (24). A previous study found that L, I, V, M, F, W, Y, and T are tolerated
250 at the C-terminus; however, this study did not test every possible amino acid at that position, and
251 our results suggest that R, K, and H may also be tolerated (24).
252
253 **Measuring IFN-γ ELISPOT responses to SIVmac239 peptides that bind to Mamu-A1*001**
254 We assessed whether the MHC peptide array can be used to efficiently screen for putative novel
255 CD8 T cell epitopes by performing IFN-γ ELISPOTs for top-binding peptides. We tested six
256 previously-established Mamu-A1*001 CD8 T cell epitopes plus the top 86 SIVmac239 peptides
257 that bound Mamu-A1*001 with the highest binding scores, using PBMCs from four SIVmac239-
258 infected Mamu-A1*001-positive rhesus macaques. PBMC from each animal ranged from 12 to 47
259 weeks post infection. Of the established epitopes, Gag CM9 was positive in all four animals; Tat
260 SL8 was positive in two (TVPWPNASL was positive in one, and LGPHYTPKIV, CAPPGYAL, and
261 CAPPGYALL were positive in none). Of the 86 peptides selected from the array, 10 were positive
262 in at least one animal and three were positive in three animals (Supplemental table II). All three
263 of the peptides positive in multiple animals (STPESANLG, STPESANLGE, TPESANLGE)
264 overlapped Tat SL8, and in one animal were positive even though SL8 itself was not. Given the
265 extensive research on Mamu-A1*001-restricted T cell responses to SIVmac239 over the past two
266 decades, it is perhaps not surprising that we did not identify additional shared responses to
267 epitopes in this virus restricted by this allele.
268
269 To quantify and confirm these responses, we produced Mamu-A1*001 tetramers folded with Gag
270 CM9, Tat SL8, and the three peptides identified by ELISPOT. We tested these for CD8 T cell

271 binding using PBMCs from two Mamu-A1*001-positive animals whose cells we used in our
272 ELISPOT experiments (one animal that did and one that did not show responses to the three
273 ELISPOT-positive peptides) and one animal that did not have Mamu-A1*001. For Gag CM9, Tat
274 SL8, STPESANLG, and STPESANLGE (but not TPESANLGE), we identified populations of
275 tetramer-positive CD8 T cells within the PBMCs from the Mamu-A1*001-positive, ELISPOT-
276 positive animal, and much smaller populations in those from the Mamu-A1*001-positive,
277 ELISPOT-negative animal (Figure 3). As expected, we did not identify populations of tetramer-
278 positive CD8 T cells in the Mamu-A1*001-negative animal.
279
280 **Comparison of Mamu-A1*001 binding peptides across SIV strains**
281 Due to its large capacity, the peptide array offers the ability to compare binding of MHC molecules
282 to peptides across several related pathogen strains. As a proof of concept, we compared Mamu-
283 A1*001 binding to SIVmac239 and SIVmac251 peptides. Most of the known Mamu-A1*001-
284 restricted SIVmac239 epitopes have the same sequence in SIVmac251, but a few have amino
285 acid variations. Of the six SIVmac239 epitopes that had amino acid variations in SIVmac251, we
286 identified one (QSPGGLDKGL in SIVmac239, QSLGGLGKGL in SIVmac251) that lost its proline
287 in position 3; as we would expect based on the importance of position 3 in binding, this epitope
288 was in the top 192 highest-binding peptides for SIVmac239 but not for SIVmac251 (Supplemental
289 table III).
290
291 We additionally compared Mamu-A1*001 binding at the region corresponding to the SIVmac239
292 Tat SL8 epitope across all viruses on the array. CD8 T cell responses to the SL8 epitope in
293 animals carrying the Mamu-A1*001 allele have been associated with suppression of viral
294 replication, and mutations in this region are commonly implicated in viral escape (26, 27).
295 Identifying whether Mamu-A1*001 is likely to be able to bind to this region despite amino acid
296 sequence differences will be important in determining whether this epitope region is likely to be

297 immunodominant in viral strains other than SIVmac239. SIVmac251 strains, also commonly used
298 in research, are split between having the SL8 sequence and TTPESANL (TL8). TL8 has
299 previously been found to be immunodominant; however, this was first established in experiments
300 studying macaques infected with SIVmac239 – suggesting that there is some latitude in Mamu-
301 A1*001's ability to bind to that region (13). The peptide array results affirm TL8 as a strong binder
302 to Mamu-A1*001, with a higher binding score (8.72) than SL8 (8.33). We found that all SIVsm
303 (isolated from sooty mangabeys) strains had the sequence PTPESANL (PL8), an escape variant
304 seen in SIVmac239 infection that exhibits reduced binding to Mamu-A1*001 (27). Similarly, on
305 our array we found that PL8 had a lower binding score than SL8 and TL8 (6.28), but one that was
306 still above baseline.
307
308 **Correlation of binding scores to IC_{50} values**
309 IC_{50} values obtained in competitive binding assays are customarily used to represent MHC-
310 peptide binding affinity; however, the equilibrium dissociation constant (k_d) is more commonly
311 used to quantify protein-ligand binding affinity and is generally preferred as it does not depend on
312 experimental conditions. The peptide array analysis outputs binding scores for each MHC-peptide
313 combination measured in arbitrary fluorescence intensity units, which reflect binding affinity. While
314 we expect these values to correlate with those obtained from competitive binding assays, recent
315 studies have found that IC_{50} values do not accurately estimate k_d values, making it challenging to
316 make a direct comparison between IC_{50} and fluorescence intensity units (28). However, using the
317 IC_{50} data available in the literature for peptides tested for binding to Mamu-A1*001, -A1*002, -
318 B*008, and -B*017, we do find that peptide array binding scores for peptides that were previously
319 found to bind an MHC molecule ($IC_{50} \leq 500$ nM) were significantly higher (Median = 0.51; 95%
320 CI = 0.34, 0.70; $p < 0.001$, using a Wilcoxon rank-sum test) than for those that did not bind MHC
321 ($IC_{50} > 500$ nM) (Figure 4). This dataset had a small selection of binding peptides for all MHC
322 molecules with non-binding peptides reported only for Mamu-B*008; using a dataset of IC_{50} values

323 for all peptide/MHC combinations generated using Stabilized Matrix Method (SMM) MHC-peptide
324 binding prediction software where experimental values were unavailable yielded the same
325 significant difference (data not shown).
326

327 ## Discussion

328 Identification of CD8 T cell epitopes has historically relied on low-throughput experiments with
329 cumbersome peptide screening methods. In an effort to develop a high-throughput method for
330 screening peptides for MHC binding prior to determining CD8 T cell responses, we used
331 ultradense peptide array technology to test the binding of class I MHC molecules to viral peptides.
332 Compared to conventional methods, ultradense peptide arrays provide massive amounts of
333 information regarding binding of class I MHC molecules to thousands of individual peptides in a
334 single experiment. In this study, of the 86 peptides that bound Mamu-A1*001 with the highest
335 fluorescence intensity scores on the array, we identified 10 peptides that were positive in at least
336 one animal and three that were positive in three animals using IFN-γ ELISPOTs and PBMCs from
337 four SIVmac239-infected Mamu-A1*001-positive rhesus macaques. These peptides were not
338 reported as epitopes previously, thus demonstrating the power of the peptide array method as an
339 alternative approach to screen for putative novel immunogenic epitopes. We additionally
340 demonstrate that many of the same epitopes previously found to bind each molecule and induce
341 CD8 T cell responses were identified as high-binding peptides in our array; 27-64% of the
342 previously-identified epitopes for each MHC molecule were found in the top 192 high-binding
343 peptides on our array.
344
345 Despite these initial promising results, many known epitopes were not detected by our method.
346 This may be due in part to the choice of 192 peptides as a convenient cutoff for "high binding."
347 Because the peptide array binding scores do not directly correspond to IC_{50} values from
348 competitive binding assays, we refrain from choosing a hard binding score cutoff to distinguish
349 binding from non-binding peptides. While there was a significant difference ($p < 0.001$) in
350 fluorescence intensity values between peptides that were previously found to bind or not bind to
351 MHC molecules based on competitive binding assays, there were some instances where

352 fluorescence intensity values did not reflect the results of the competitive binding assays. For
353 example,the Tat SL8 epitope's high fluorescence intensity (7.15 sd above the median for Mamu-
354 A1*001) and low IC_{50} value both indicate strong peptide binding. However, the Gag CM9 epitope
355 showed a substantially lower fluorescence intensity (2.36 sd above the median for Mamu-
356 A1*001), but also had a low IC_{50} value. These discrepancies may be attributable to differences in
357 how certain peptides or MHC molecules perform on the array; certainly, it would be reasonable
358 to expect that some differences may arise because peptides are anchored on the array rather
359 than free-floating. Unfortunately, neither competitive binding assays nor peptide arrays truly mimic
360 *in vivo* MHC-peptide binding, so without further optimization of the method, it would be premature
361 to label one method as "more true" than the other.
362
363 We tested a portion of the high-binding peptides identified here by IFN-γ ELISPOT as a proof of
364 concept. We identified three SIVmac239 peptides that strongly bound Mamu-A1*001 that had
365 positive responses in at least two animals and were not previously identified epitopes; two of the
366 tetramers generated to test these responses were functional. All three peptides overlapped the
367 Tat SL8 epitope, and the two functional tetramers had weaker responses than the SL8 tetramer,
368 which may suggest that these are not separate epitopes but simply extensions of the SL8 minimal
369 optimal epitope. Still, this indicates that a researcher using the array to look for MHC binding to
370 an unknown pathogen's proteome could reasonably expect to find CD8 T cell epitopes in a subset
371 of high-binding peptides identified by the array; moreover, this demonstrates that, at least in some
372 cases, peptide binding to clusters of related sequences (ie., 8-mer, 9-mer, and 10-mer peptides
373 containing the same component 8-mer) could be used as a filter to identify interesting peptides
374 for follow-up. Using the array as a screening tool could thus drastically reduce the amount of work
375 and resources required to identify novel epitopes.
376

377 MHC-peptide array technology has not yet been optimized to work with class II MHC molecules.
378 CD8 T cell responses are not the sole arbiter of protective immunity in HIV vaccination; a better
379 understanding of CD4 responses is necessary for developing a more comprehensive picture of
380 the determinants of a successful vaccine. Assessing CD4 T cell responses in PBMCs and
381 developing class II tetramers is currently more difficult than for CD8 T cells and class I MHC.
382 While this will present challenges in validating a class II array, it underscores the need for a
383 method for rapidly screening for potential CD4 T cell responses, and optimizing a class II array
384 will be an important future direction.
385
386 Ultradense peptide arrays offer a method for rapidly mapping class I MHC binding to thousands
387 of individual peptides. While we used peptides derived from SIV and SHIV here, the scalability of
388 ultradense peptide arrays makes them an attractive option for characterizing MHC class I binding
389 to peptides from larger, more complex pathogens. *Mycobacterium tuberculosis* (Mtb) Erdman
390 strain, for instance, has 4,246 annotated protein sequences, in contrast to nine for SIV. The size
391 of the Mtb proteome makes thorough analyses of MHC binding across the full proteome nearly
392 impossible with conventional methods, but just three peptide array chips would be sufficient to
393 accommodate a full 8-, 9-, and 10-mer peptide analysis for a single MHC allele. The ability to
394 screen peptides for MHC binding en masse prior to performing *in vitro* tests will streamline epitope
395 identification studies that might otherwise be prohibitively large and expensive to conduct.
396
397 **Acknowledgments**
398 The authors would like to thank Amy Ellis for her help in analyzing the tetramer flow data.
399
400

401 ## References

402 1. Hewitt, E. W. 2003. The MHC class I antigen presentation pathway: strategies for viral
403 immune evasion. *Immunology* 110: 163-169.
404 2. Halle, S., O. Halle, and R. Förster. 2017. Mechanisms and Dynamics of T Cell-Mediated
405 Cytotoxicity In Vivo. *Trends Immunol* 38: 432-443.
406 3. Wieczorek, M., E. T. Abualrous, J. Sticht, M. Álvaro-Benito, S. Stolzenberg, F. Noé, and C.
407 Freund. 2017. Major Histocompatibility Complex (MHC) Class I and MHC Class II
408 Proteins: Conformational Plasticity in Antigen Presentation. *Front Immunol* 8: 292.
409 4. McMichael, A. J., and W. C. Koff. 2014. Vaccines that stimulate T cell immunity to HIV-1:
410 the next step. *Nat Immunol* 15: 319-322.
411 5. Zhao, L., M. Zhang, and H. Cong. 2013. Advances in the study of HLA-restricted epitope
412 vaccines. *Hum Vaccin Immunother* 9: 2566-2577.
413 6. Peters, B., H. H. Bui, J. Sidney, Z. Weng, J. T. Loffredo, D. I. Watkins, B. R. Mothé, and A.
414 Sette. 2005. A computational resource for the prediction of peptide binding to Indian
415 rhesus macaque MHC class I molecules. *Vaccine* 23: 5212-5224.
416 7. Di Carluccio, A. R., C. F. Triffon, and W. Chen. 2018. Perpetual complexity: predicting
417 human CD8. *Immunol Cell Biol* 96: 358-369.
418 8. E Silva, R. F., L. F. Ferreira, M. Z. Hernandes, M. E. de Brito, B. C. de Oliveira, A. A. da
419 Silva, O. P. de-Melo-Neto, A. M. Rezende, and V. R. Pereira. 2016. Combination of In
420 Silico Methods in the Search for Potential CD4(+) and CD8(+) T Cell Epitopes in the
421 Proteome of Leishmania braziliensis. *Front Immunol* 7: 327.
422 9. Mahdavi, M., and V. Moreau. 2016. In silico designing breast cancer peptide vaccine for
423 binding to MHC class I and II: A molecular docking study. *Comput Biol Chem* 65: 110-116.
424 10. Budde, M. L., J. J. Lhost, B. J. Burwitz, E. A. Becker, C. M. Burns, S. L. O'Connor, J. A.
425 Karl, R. W. Wiseman, B. N. Bimber, G. L. Zhang, W. Hildebrand, V. Brusic, and D. H.

426 O'Connor. 2011. Transcriptionally abundant major histocompatibility complex class I
427 alleles are fundamental to nonhuman primate simian immunodeficiency virus-specific
428 CD8+ T cell responses. *J Virol* 85: 3250-3261.
429 11. Sims, S., C. Willberg, and P. Klenerman. 2010. MHC-peptide tetramers for the analysis of
430 antigen-specific T cells. *Expert Rev Vaccines* 9: 765-774.
431 12. Fiore-Gartland, A., B. A. Manso, D. P. Friedrich, E. E. Gabriel, G. Finak, Z. Moodie, T.
432 Hertz, S. C. De Rosa, N. Frahm, P. B. Gilbert, and M. J. McElrath. 2016. Pooled-Peptide
433 Epitope Mapping Strategies Are Efficient and Highly Sensitive: An Evaluation of Methods
434 for Identifying Human T Cell Epitope Specificities in Large-Scale HIV Vaccine Efficacy
435 Trials. *PLoS One* 11: e0147812.
436 13. Allen, T. M., B. R. Mothé, J. Sidney, P. Jing, J. L. Dzuris, M. E. Liebl, T. U. Vogel, D. H.
437 O'Connor, X. Wang, M. C. Wussow, J. A. Thomson, J. D. Altman, D. I. Watkins, and A.
438 Sette. 2001. CD8(+) lymphocytes from simian immunodeficiency virus-infected rhesus
439 macaques recognize 14 different epitopes bound by the major histocompatibility complex
440 class I molecule mamu-A*01: implications for vaccine design and testing. *J Virol* 75: 738-
441 749.
442 14. Mothé, B. R., J. Sidney, J. L. Dzuris, M. E. Liebl, S. Fuenger, D. I. Watkins, and A. Sette.
443 2002. Characterization of the peptide-binding specificity of Mamu-B*17 and identification
444 of Mamu-B*17-restricted epitopes derived from simian immunodeficiency virus proteins. *J*
445 *Immunol* 169: 210-219.
446 15. Loffredo, J. T., J. Sidney, C. Wojewoda, E. Dodds, M. R. Reynolds, G. Napoé, B. R.
447 Mothé, D. H. O'Connor, N. A. Wilson, D. I. Watkins, and A. Sette. 2004. Identification of
448 seventeen new simian immunodeficiency virus-derived CD8+ T cell epitopes restricted by
449 the high frequency molecule, Mamu-A*02, and potential escape from CTL recognition. *J*
450 *Immunol* 173: 5064-5076.

451 16. Loffredo, J. T., J. Sidney, A. T. Bean, D. R. Beal, W. Bardet, A. Wahl, O. E. Hawkins, S.
452 Piaskowski, N. A. Wilson, W. H. Hildebrand, D. I. Watkins, and A. Sette. 2009. Two MHC
453 class I molecules associated with elite control of immunodeficiency virus replication,
454 Mamu-B*08 and HLA-B*2705, bind peptides with sequence similarity. *J Immunol* 182:
455 7763-7775.
456 17. John T. Loffredo, L. E. V., David I. Watkins. 2006. Beyond Mamu-A*01+ Indian Rhesus
457 Macaques: Continued Discovery of New MHC Class I Molecules that Bind Epitopes from
458 he Simian AIDS Viruses. *HIV Molecular Immunology* 29-51.
459 18. Martins, M. A., D. C. Tully, M. A. Cruz, K. A. Power, M. G. Veloso de Santana, D. J. Bean,
460 C. B. Ogilvie, R. Gadgil, N. S. Lima, D. M. Magnani, K. Ejima, D. B. Allison, M. Piatak, J.
461 D. Altman, C. L. Parks, E. G. Rakasz, S. Capuano, R. Galler, M. C. Bonaldo, J. D. Lifson,
462 T. M. Allen, and D. I. Watkins. 2015. Vaccine-Induced Simian Immunodeficiency Virus-
463 Specific CD8+ T-Cell Responses Focused on a Single Nef Epitope Select for Escape
464 Variants Shortly after Infection. *J Virol* 89: 10802-10820.
465 19. Heffron, A. S., E. L. Mohr, D. Baker, A. K. Haj, C. R. Buechler, A. Bailey, D. M. Dudley, C.
466 M. Newman, M. S. Mohns, M. Koenig, M. E. Breitbach, M. Rasheed, L. M. Stewart, J.
467 Eickhoff, R. S. Pinapati, E. Beckman, H. Li, J. Patel, J. C. Tan, and D. H. O'Connor. 2018.
468 Antibody responses to Zika virus proteins in pregnant and non-pregnant macaques. *PLoS*
469 *Negl Trop Dis* 12: e0006903.
470 20. Mishra, N., A. Caciula, A. Price, R. Thakkar, J. Ng, L. V. Chauhan, K. Jain, X. Che, D. A.
471 Espinosa, M. Montoya Cruz, A. Balmaseda, E. H. Sullivan, J. J. Patel, R. G. Jarman, J. L.
472 Rakeman, C. T. Egan, C. B. E. M. Reusken, M. P. G. Koopmans, E. Harris, R. Tokarz, T.
473 Briese, and W. I. Lipkin. 2018. Diagnosis of Zika Virus Infection by Peptide Array and
474 Enzyme-Linked Immunosorbent Assay. *MBio* 9:
475 21. Hutchinson, S. L., L. Wooldridge, S. Tafuro, B. Laugel, M. Glick, J. M. Boulter, B. K.
476 Jakobsen, D. A. Price, and A. K. Sewell. 2003. The CD8 T cell coreceptor exhibits

477 disproportionate biological activity at extremely low binding affinities. *J Biol Chem* 278:
478 24285-24293.
479 22. Forsström, B., B. B. Axnäs, K. P. Stengele, J. Bühler, T. J. Albert, T. A. Richmond, F. J.
480 Hu, P. Nilsson, E. P. Hudson, J. Rockberg, and M. Uhlen. 2014. Proteome-wide epitope
481 mapping of antibodies using ultra-dense peptide arrays. *Mol Cell Proteomics* 13: 1585-
482 1597.
483 23. Reynolds, M. R., A. M. Weiler, S. M. Piaskowski, M. Piatak, H. T. Robertson, D. B. Allison,
484 A. J. Bett, D. R. Casimiro, J. W. Shiver, N. A. Wilson, J. D. Lifson, W. C. Koff, and D. I.
485 Watkins. 2012. A trivalent recombinant Ad5 gag/pol/nef vaccine fails to protect rhesus
486 macaques from infection or control virus replication after a limiting-dose heterologous SIV
487 challenge. *Vaccine* 30: 4465-4475.
488 24. Sidney, J., J. L. Dzuris, M. J. Newman, R. P. Johnson, A. Kaur, K. Amitinder, C. M.
489 Walker, E. Appella, B. Mothe, D. I. Watkins, and A. Sette. 2000. Definition of the Mamu
490 A*01 peptide binding specificity: application to the identification of wild-type and optimized
491 ligands from simian immunodeficiency virus regulatory proteins. *J Immunol* 165: 6387-
492 6399.
493 25. Allen, T. M., J. Sidney, M. F. del Guercio, R. L. Glickman, G. L. Lensmeyer, D. A. Wiebe,
494 R. DeMars, C. D. Pauza, R. P. Johnson, A. Sette, and D. I. Watkins. 1998.
495 Characterization of the peptide binding motif of a rhesus MHC class I molecule (Mamu-
496 A*01) that binds an immunodominant CTL epitope from simian immunodeficiency virus. *J*
497 *Immunol* 160: 6062-6071.
498 26. Loffredo, J. T., B. J. Burwitz, E. G. Rakasz, S. P. Spencer, J. J. Stephany, J. P. Vela, S. R.
499 Martin, J. Reed, S. M. Piaskowski, J. Furlott, K. L. Weisgrau, D. S. Rodrigues, T. Soma, G.
500 Napoé, T. C. Friedrich, N. A. Wilson, E. G. Kallas, and D. I. Watkins. 2007. The antiviral
501 efficacy of simian immunodeficiency virus-specific CD8+ T cells is unrelated to epitope
502 specificity and is abrogated by viral escape. *J Virol* 81: 2624-2634.

503 27. Allen, T. M., D. H. O'Connor, P. Jing, J. L. Dzuris, B. R. Mothé, T. U. Vogel, E. Dunphy, M.
504 E. Liebl, C. Emerson, N. Wilson, K. J. Kunstman, X. Wang, D. B. Allison, A. L. Hughes, R.
505 C. Desrosiers, J. D. Altman, S. M. Wolinsky, A. Sette, and D. I. Watkins. 2000. Tat-specific
506 cytotoxic T lymphocytes select for SIV escape variants during resolution of primary
507 viraemia. *Nature* 407: 386-390.
508 28. Barbet, J., and S. Huclier-Markai. 2019. Equilibrium, affinity, dissociation constants, IC5O:
509 Facts and fantasies. *Pharm Stat*
510
511

512 **Table I:** The percentage of known epitopes for each MHC molecule that were found in the top N
513 peptides with high signal intensity.
514
515 **Figure 1: MHC-peptide binding plots.** Previously-identified SIVmac239 8-mer, 9-mer, and 10-
516 mer epitopes found in the top 192 peptides for each MHC class I molecule. Parentheticals
517 indicate overlapping epitopes.
518
519 **Figure 2: Identification of anchor residues by substitution analysis.** Single amino acid
520 substitutions were made at each position of the Gag CM9 (A) and SL8 (B) epitopes. Each box
521 represents a peptide with the indicated substitution, and the fold change in signal intensities for
522 each substitution (compared to the wild-type sequence) is reported. Position numbers are
523 reported along the left axis of the heatmap; the fold change color scale is indicated at right.
524
525 **Figure 3: Tetramer analysis.** Mamu-A1*001 MHC-peptide tetramers were generated and
526 tested for binding to PBMC from SIV-infected animals. (A) A representative sequential gating
527 strategy for the flow cytometry data is shown. (B) Tetramer binding to PBMC from three animals
528 was tested; shown are the ELISPOT responses to each peptide (STPESANLG, STPESANLGE,
529 and TPESANLG are abbreviated as SG9, SE10, and TE9 respectively), and the percentage of
530 CD8 T cells that were tetramer-positive.
531
532 **Figure 4: Comparison of IC_{50} and fluorescence intensity scores.** Peptides with
533 experimentally obtained IC_{50} values for binding to Mamu-A1*001, -A1*002, -B*008, or -B*017
534 were grouped by IC_{50} value into binding ($IC_{50} \leq 500$ nM) and non-binding ($IC_{50} > 500$ nM)
535 peptides. Binding peptides had significantly higher fluorescence intensity scores than non-
536 binding peptides ($p < 0.001$, using a Wilcoxon rank-sum test).

	N = 24	N = 48	N = 96	N = 192
Mamu-A1*001	22%	44%	47%	64%
Mamu-A1*002	0%	0%	16%	37%
Mamu-B*008	1%	8%	12%	27%
Mamu-B*017	12%	18%	29%	41%
Average coverage	9%	18%	26%	42%

Table I

Figure 1

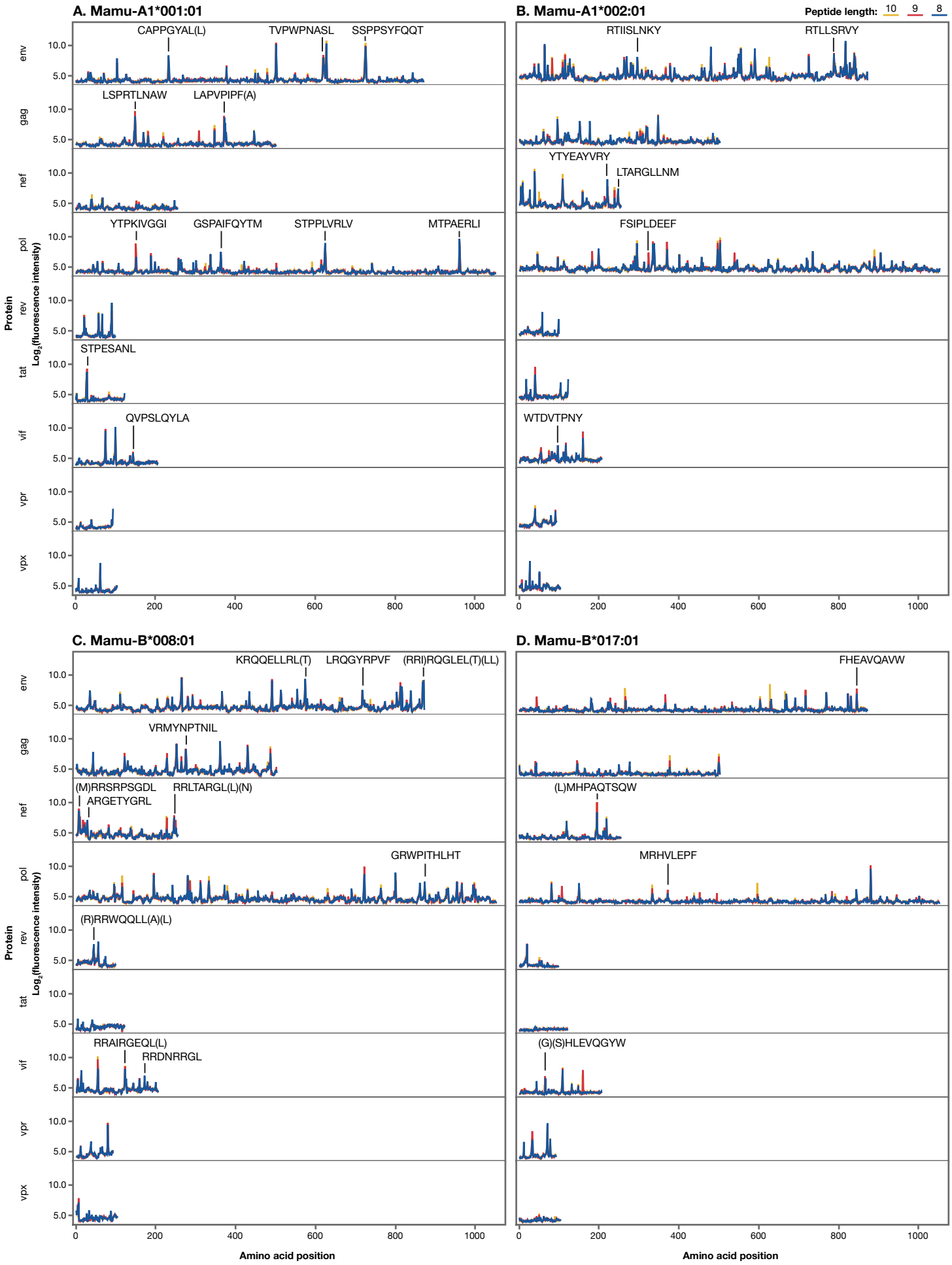

A. Mamu-A1*001:01
B. Mamu-A1*002:01
Peptide length: 10 9 8
CAPPGYAL(L)
TVPWPNASL
SSPPSYFQQT
LSPRTLNAW
LAPVPIPF(A)
YTPKIVGGI
GSPAIFQYTM
STPPLVRLV
MTPAERLI
STPESANL
QVPSLQYLA
RTIISLNKY
RTLLSRVY
YTYEAYVRY
LTARGLLNM
FSIPLDEEF
WTDVTPNY
C. Mamu-B*008:01
D. Mamu-B*017:01
KRQQELLRL(T)
LRQGYRPVF
(RRI)RQGLEL(T)(LL)
VRMYNPTNIL
(M)RRSRPSGDL
ARGETYGRL
RRLTARGL(L)(N)
GRWPITHLHT
(R)RRWQQLL(A)(L)
RRAIRGEQL(L)
RRDNRRGL
FHEAVQAVW
(L)MHPAQTSQW
MRHVLEPF
(G)(S)HLEVQGYW
Protein
env
gag
nef
pol
rev
tat
vif
vpr
vpx
Log2(fluorescence intensity)
10.0
5.0
0
200
400
600
800
1000
Amino acid position

Figure 2

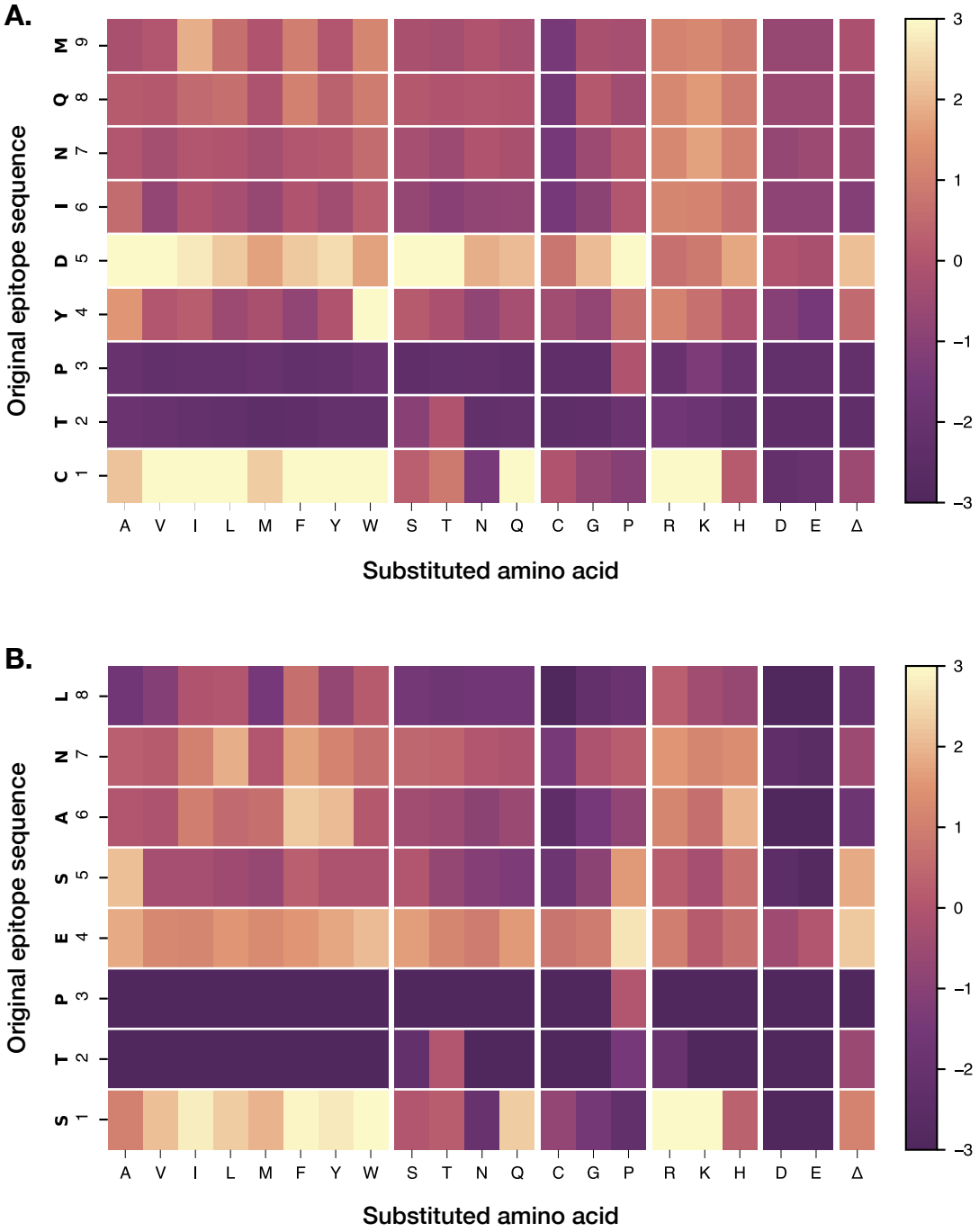
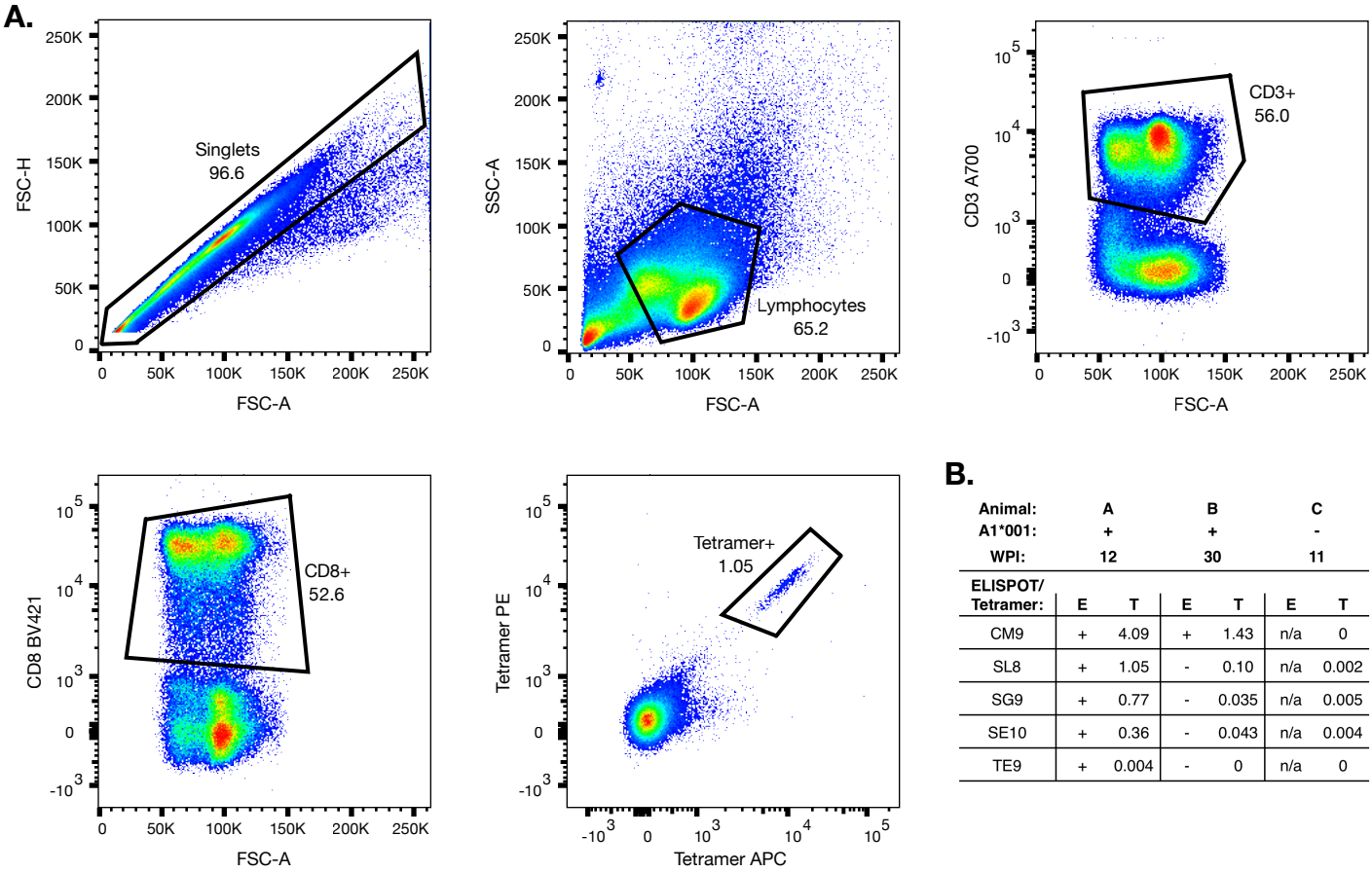

A.
Original epitope sequence
M 9
Q 8
N 7
I 6
D 5
Y 4
P 3
T 2
C 1
A V I L M F Y W S T N Q C G P R K H D E Δ
Substituted amino acid
3
2
1
0
−1
−2
−3
B.
Original epitope sequence
L 8
N 7
A 6
S 5
E 4
P 3
T 2
S 1
A V I L M F Y W S T N Q C G P R K H D E Δ
Substituted amino acid
3
2
1
0
−1
−2
−3

Figure 3

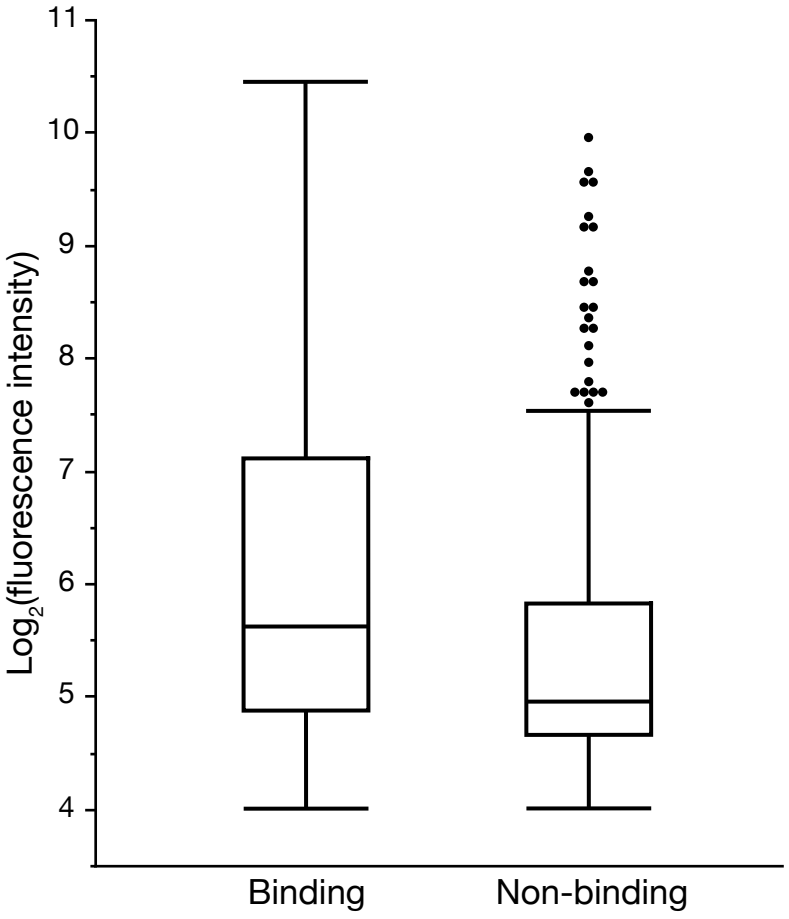
A.



B.

Animal:	A		B		C	
A1*001:	+		+		-	
WPI:	12		30		11	
ELISPOT/ Tetramer:	E	T	E	T	E	T
CM9	+	4.09	+	1.43	n/a	0
SL8	+	1.05	-	0.10	n/a	0.002
SG9	+	0.77	-	0.035	n/a	0.005
SE10	+	0.36	-	0.043	n/a	0.004
TE9	+	0.004	-	0	n/a	0

Figure 4


11
10
9
8
7
6
5
4
Log2(fluorescence intensity)
Binding
Non-binding