Most viral peptides displayed by class I MHC on infected cells are immunogenic

Nathan P. Croft, b, 1, Stewart A. Smith, c Jana Pickering, d John Sidney, e Bjorn Peters, e, f Pouya Faridi, a, b, Matthew J. Witney, c Prince Sebastian, c Inge E. A. Flesch, c Sally L. Heading, c Alessandro Sette, e, f Nicole L. La Gruta, a, b, f Anthony W. Purcell, a, b, 1, 2, and David C. Tscharke c, d, e

Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia; bDepartment of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia; cJohn Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; dDivision of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037; eDepartment of Medicine, University of California, San Diego, La Jolla, CA 92039; and fDepartment of Microbiology and Immunology, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia

Edited by Emil R. Unanue, Washington University, St. Louis, MO, and approved January 4, 2019 (received for review September 4, 2018)

CD8+ T cells are essential effectors in antiviral immunity, recognizing short virus-derived peptides presented by MHC class I (pMHC) on the surface of infected cells. However, the fraction of viral pMHC on infected cells that are immunogenic has not been shown for any virus. To address this fundamental question, we used peptide sequencing by high-resolution mass spectrometry to identify more than 170 vaccinia virus pMHC presented on infected mouse cells. Next, we screened each peptide for immunogenicity in multiple virus-infected mice, revealing a wide range of immunogenicities. A surprisingly high fraction (>80%) of pMHC were immunogenic in at least one infected mouse, and nearly 40% were immunogenic across more than half of the mice screened. The high number of peptides found to be immunogenic and the distribution of responses across mice give us insight into the specificity of antiviral CD8+ T cell responses.

Results and Discussion

Comprehensive Identification of VACV-Derived pMHC. To generate a comprehensive list of the VACV pMHCIs that might be presented during infection of C57BL/6 mice, we used DC2.4 cells as the substrate because they have high expression levels of H-2Kb haplotype MHCI and high infection efficiency with VACV (12). The cells were infected with VACV for 6 h because we have demonstrated previously that presentation of epitopes from all kinetic classes can be detected at this time (12, 13). H-2Kb and Db molecules were individually isolated by immunoaffinity purification, and acid-eluted peptides were then analyzed by high-resolution liquid chromatography–tandem mass spectrometry (LC-MS/MS). Spectra were searched against the VACV proteome. The distribution of the candidate VACV peptides according to their statistical confidence shows a clustering of sequences at the top, with around half having a confidence of >95% (Fig. 1A). We applied an inclusive threshold of 50% based on the appearance of two previously identified epitopes with confidence levels as low as 70%. Subsequent screening of the same data against combined mouse and VACV proteomes flagged several ambiguous spectra that could be assigned as either a VACV- or mouse-derived sequence, resulting in a list of 191 peptides (Dataset S1). Next, we validated this list in two ways: First, the integrity of the MS data were tested by either assessing the degree of similarity between spectra from peptides eluted from infected cells and that of a corresponding synthetic peptide (139 peptides) or through the sensitive detection method of multiple reaction monitoring (MRM) of mock versus infected virus.

Significance

CD8+ T cells are key to the defense of animals against virus infection. These immune cells recognize peptides derived from viral proteins that are displayed on the surface of infected cells in a complex with host proteins known as MHC I. Many viral peptides are displayed by MHC I on infected cells, but it has never been shown what fraction of these can induce an immune response. We answered this long-standing question, finding that more than 80% of vaccinia virus peptides presented by MHC I on infected mouse cells were immunogenic across a population of mice.

Author contributions: N.P.C., N.L.L.G., A.W.P., and D.C.T. designed research; N.P.C., S.A.S., J.P., J.S., M.I.W., P.S., I.E.A.F., and S.L.H. performed research; N.P.C., J.S., B.P., P.F., and D.C.T. analyzed data; and N.P.C., J.S., B.P., P.F., A.S., N.L.L.G., A.W.P., and D.C.T. wrote the paper.

The authors declare no conflict of interest.

This is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: LC-MS/MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010811.

1To whom correspondence may be addressed. Email: nathan.croft@monash.edu, anthony.purcell@monash.edu, or david.tscharke@anu.edu.au.

2A.W.P. and D.C.T. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1815239116/-/DCSupplemental.

Published online February 4, 2019.
and D\textsuperscript{\textalpha} (SI Appendix, Fig. S1 B and C). Taking this a step further we measured the binding of these peptides to MHCI, which we note correlated very well with predictions made with NetMHC 4.1 (SI Appendix, Fig. S1D), and demonstrated that 84% of the VACV-derived peptides bound MHCI with <500 nM affinity (Fig. 1E). Finally, we note that our rate of discovery (73 and 97 for H-2K\textalpha\textbeta and D\textalpha, respectively) was similar to that published for VACV in the context of two human MHC allomorphs (110 and 64 peptides for HLA-A2 and B7, respectively) (17).

Characteristics of VACV-Derived, MHCI-Presented Peptides. The 172 peptides identified by LC-MS/MS were spread across the VACV proteome, being encoded by the four kinetic classes of viral genes (18–20). Similar to other recent studies (4, 6, 17, 21), we found at least one epitope in 40% of all VACV proteins (92 of 230 proteins), but some of these proteins were richer sources of presented peptides (SI Appendix, Fig. S2A). In some proteins, the multiple pMHCI detected comprised sets of peptides that were largely overlapping but with extensions at either the amino- or carboxy-terminus (for example, proteins J4 and F5, respectively; Dataset S1). In other cases, there were multiple non-overlapping peptides from the same protein. Most of the proteins here that are sources of multiple, nonoverlapping pMHCI are immunoprevalent, having multiple epitopes presented in the context of several MHCI allomorphs, e.g., A10, A3, A17, B8, and J6 as defined by others (6, 17). Others are unique to this study, e.g., A18, with five peptides and A23, A24, and A8, with three peptides each. There is no obvious characteristic that links these proteins, but understanding why they are so frequently presented should help open one of the remaining black boxes in antigen presentation: The selection of antigens and peptide precursors that occurs before loading onto MHCI. To understand the frequency of presentation across the proteome in a way that takes into account the different sizes of viral proteins, we calculated the number of presented peptides per 100 aa for each protein. These proteins were then grouped according to kinetic class, allowing an analysis of the role of expression time on pMHCI display that was not confounded by the differing numbers and sizes of genes in each class (Fig. 2B). This showed that there was no significant difference in the frequency of epitope occurrence by any kinetic class, suggesting that all are equally available to the antigen presentation pathway (Fig. 2B). To extend this analysis, we analyzed the abundance of pMHCI as inferred by the precursor ion intensity and the number of times a given peptide was identified by the mass spectrometer (i.e., spectral counts). These methods are approximations, but we note that they broadly correlate with each other across our data (SI Appendix, Fig. S2B). Intriguingly, the earliest class of genes was found to be the source of significantly more pMHCI than the later classes when normalized to their coding capacity (Fig. 2 C and D), but this was not related to the affinity of these peptides for MHCI (SI Appendix, Fig. S2C). Looking at the individual protein level, the distribution of presented peptides within a viral protein shows that MHCI ligands are equally likely to be processed from any part of an antigen (SI Appendix, Fig. S2D). The start positions of the peptides were distributed along the entire length of the protein, with the average sitting at almost the exact midpoint of the protein, both for H-2K\textalpha\textbeta and for D\textalpha-binding peptides. Finally, VACV has been the subject of several studies to identify MHCI-restricted epitopes and is already the best characterized large virus in this regard, with >70 VACV-derived, H-2\textsuperscript{\textalpha}–restricted peptides or epitopes identified (Immune Epitope Database, IEDB; ref. 22). Most of these peptides were found first or confirmed in a single, landmark, proteome-wide study that identified 49 epitopes from a library of ~2,500 predicted pMHCI (4). This study set the benchmark for identification of CD8+ T cell epitopes by prediction and screening and provides a good dataset for comparison with our LC-MS/MS identified peptides. The previously identified epitopes only partially overlapped with the MHCI ligands identified here (Fig. 2E). This suggests limitations in all approaches. For our LC-MS/MS study,
The peptides T Cell Epitopes. The protein sources of a viral immunogenicity screen to the 46 VACV-derived H-2b-restricted peptides/epitopes from the IEDB that were not found by LC-MS/MS, in addition to one entirely unpublished and one more finely mapped version of a published (26) epitope from other work in our laboratories. Combining these with the set found by LC-MS/MS gave 220 VACV peptides tested eight times, giving a number of times positive and the average size of response for each peptide. As expected from the above, the average size of the CD8+ T cell response across this complete set of peptides correlated with the frequency with which a response was detected (SI Appendix, Fig. S4A). This correlation was significant ($r = 0.8219, P < 0.0001$) by a Spearman rank test. Across the 220 pMHCI in this extended set, 84 were defined as major epitopes, and most of these (67 peptides) were found by our LC-MS/MS experiment (SI Appendix, Fig. S4B). By contrast, the large number of tests, from nil to eight. However, an examination of the distribution of responses across the eight tests suggests a more nuanced picture. In general, individual responses are distributed evenly around a mean that rises for peptides that are found to be immunogenic in a greater number of mice. This can be seen for selected examples (Fig. 3A) and also when the data are plotted in aggregate for all mice and peptides (SI Appendix, Fig. S3A). Indeed, the increase in population mean as pMHCI are positive in more than half of the mice (five to eight of eight mice); and minor epitopes, which are the remaining pMHCI that were less frequently positive (one to four of eight mice). Perhaps this occurs because particular T cell receptors are not available in all mice. Further, we note all assays for immunogenicity are subject to a threshold of detection. Altogether, we conclude that from an experimental point of view, immunogenicity will typically be found to be broader across a population than within a single individual, but in most cases this will reflect limits of detection and not absolute immunogenicity. As a final note, we tested mice by i.p. infection here, but immunogenicity of VACV epitopes can be influenced by route of infection, dose, or virulence of the strain used (24, 25).

The use of thresholds in individual mice also enables pMHCI to be divided into immunogenicity groups for the purposes of further analyses. We have divided the pMHCI into three groups: Nonimmunogenic, which were never positive; major epitopes, which were positive in more than half of the mice (five to eight of eight mice); and minor epitopes, which are the remaining pMHCI that were less frequently positive (one to four of eight mice). Using this definition, we found 39 and 44% of pMHCI detected on VACV-infected cells to be major and minor epitopes, respectively (Fig. 3B).

A Comprehensive Set of VACV CD8+ T Cell Epitopes. The peptides detected by LC-MS/MS only partially overlapped the known H-2b-restricted epitopes for VACV. So we extended the immunogenicity screen to the 46 VACV-derived H-2b-restricted peptides/epitopes from the IEDB that were not found by LC-MS/MS. 0.8219, $P < 0.0001$) by a Spearman rank test. Across the 220 pMHCI in this extended set, 84 were defined as major epitopes, and most of these (67 peptides) were found by our LC-MS/MS experiment (SI Appendix, Fig. S4B). By contrast, the large
and D cells (Fig. 4). As a readout, this allowed us to determine the fraction of the total response contributed by specificities found by LC-MS/MS here (SI Appendix, Fig. S4C and D). We note that the original predictive study was done over a decade ago, and significant advances in HLA-binding predictions have occurred since then, which we expect will improve their accuracy. Our data will be useful to objectively evaluate the performance of newer algorithms and to identify weaknesses and refine HLA-binding predictions (e.g., as in SI Appendix, Fig. S1D).

Summing the responses to all nonoverlapping major and minor epitopes shows that the vast majority of anti-VACV CD8+ T cells are elicited by the major epitopes (Fig. 3D). Next, to estimate the fraction of the total response to VACV that is contributed to VACV-specific T cells, we compared the fraction of CD8+ T cells from spleens of acutely infected mice that were CD62L+ and Granzyme B+ with the summed response to the 90 most immunogenic peptides (excluding overlapping peptides as noted above) in these same mice (Fig. 3E). It is not possible to test all of the peptides in a single mouse, but the 90 most immunogenic peptides account for 95% of the response found by all of the immunogenic peptides, so is a close estimate of the total now mapped. Using CD62L and Granzyme B is a more accurate way to determine the total anti-VACV CD8+ response than using virus-infected cells as antigen-presenting cells, as it is a readout of intracellular IFN-γ, as the latter method underestimates the response by around 50% (10). The comparison of these measures finds that these 90 pMHC account for around 70% of the total anti-VACV CD8+ T cell response (Fig. 3E). This suggests that despite the depth of all of the studies to date, further specificities remain to be defined. In addition to the limitations of using DC2.4 as the substrate for LC-MS/MS, there may also be immunogenic peptides that have post-translational modifications that we did not test, or come from outside conventional ORFs or that are spliced (27–31). Finally, the 25 most immunogenic peptides are shown in a hierarchy (SI Appendix, Fig. S4E), excluding any less immunogenic variants from overlapping sets. This hierarchy includes four epitopes that we have refined by identifying more immunogenic variants here (e.g., A189-198 ITYRFYLINL, which is almost twice as immunogenic as the previously defined A189-196 ITYRFYLI), and eight peptides for which no variants exist in the IEDB.

Factors Contributing to Immunogenicity. Noting the importance of the major CD8+ T cell epitopes in the total response, we then asked what factors were associated with these larger and more prevalent responses, examining time of expression, amount presented, and affinity of peptides for MHC.

First, the distribution of anti-VACV responses according to antigen and arranged by kinetic class visually suggests that late viral proteins were recognized by fewer CD8+ T cells (Fig. 4A) as previously suggested (32). We examined this more quantitatively in the same way that we analyzed presentation on MHC, by calculating the occurrence of major epitopes (Fig. 4B) and the total size of the response (Fig. 4C) per 100 aa for each viral protein. This allowed an unbiased comparison of immunogenicity across the kinetic classes. Immunogenicity was not significantly different across the kinetic classes of genes as a whole, and the distribution of major epitopes was not significantly different across the kinetic classes ($P = 0.2353$, Fisher’s exact test). However, the distribution of highly immunogenic proteins, defined as the top 10% by total size of
response, was significantly different \((P = 0.04033, \text{Fisher's exact test})\), with no late proteins in this category. Overall then, our data support the notion that while late genes are equally likely to be a source of pMHC on infected cells (Fig. 2), their epitopes are less likely to be highly immunogenic.

Second, we investigated epitope abundance as inferred from precursor ion intensities (Fig. 4D) and spectral counts (Fig. 4E). On average, major pMHC were identified by the mass spectrometer significantly more than minor, or nonimmunogenic peptides. A similar trend was seen with precursor ion abundance, but pair-wise comparisons between groups were not statistically significant. Together these analyses suggest that abundance plays some role in determining immunogenicity, but the evidence was not especially strong. This conclusion is supported by the finding that the E1.1 class of VACV genes was more abundantly presented (Fig. 2 C and D), but not more immunogenic \((P < 0.0001 \text{ and } P = 0.0005, \text{respectively, Kruskal–Wallis})\). While at face value this seems to be an obvious finding, we emphasize that this analysis only included peptides that were eluted from MHCII, and so all are of sufficient affinity to be presented. Further, we ruled out a role for confidence in the assignement of peptide identity in the original LC-MS/MS data as a factor that may contribute to the observed correlates with immunogenicity \((SI \ \text{Appendix, Fig. S5 A and B})\).

Across these three factors, affinity for MHC was the most rigorously associated with dominance, even in the set of peptides we identified as being presented on infected cells. Therefore, we speculate that the affinity of a peptide for MHC may have a significance for CD8+ T cell immunity beyond making a pMHC available at the cell surface.

**Summary.** To conclude, this study provides the largest resource of LC-MS/MS-verified MHCII ligands along with their immunogenicity for any viral infection model. Using this information it was found that >80% of viral pMHC can be immunogenic across a set of individuals, a surprisingly high fraction. Immunogenicity was found to be higher on a population basis than in individuals due to weak epitopes not stimulating responses above the limit of detection in all individuals. Finally, we found that while roles can be shown for abundance and time of gene expression, major CD8+ T cell responses are most strongly associated with high affinity for MHCII, even among peptides that are known to be presented.

**Materials and Methods**

Detailed methods are provided in SI Appendix.

**Viruses and Cell Lines.** Surfocote purified stocks VACV strain Western Reserve (ATCC #VR1354) and DC2.4 cells (33) were used for LC-MS/MS experiments.

**Infection of Cells for LC-MS/MS.** Many \(1 \times 10^7\) DC2.4 cells \((5 \times 10^7 \text{ cells for MRM validations})\) were infected at 10 plaque-forming units (PFU) per cell for a total of 6 h at 37 °C, after which cells were centrifuged, media removed, and the pellet snap frozen.

**Mice and Infections.** Specific pathogen-free female C57BL/6 mice 8–16 wk of age were obtained from The Australian National University (ANU) Phenomics Facility (Canberra, Australia). Mice were housed, and experiments were done according to the relevant ethical requirements and under an approval from the ANU animal ethics and experimentation committee \((\text{approvals A2011.01, A2013.37, and A2016.45})\). Mice were infected by the i.p. route with \(1 \times 10^7\) PFU of VACV in 200 μL of PBS. To test all peptides eight times, they were screened in two main batches, so 16 mice were required for the main screen. Additional mice were then used to fill data for some failed assays in the main screen.

**Identification of pMHC by LC-MS/MS.** VACV-infected cells were lysed, and pMHC complexes were purified \((12)\). Briefly, cell lysates were cleared by centrifugation, and pMHC immunooaffinity was purified on Sepharose beads with anti-Kb or D0 before elution with 10% acetic acid. The mixture of peptides and MHC protein chains was fractionated on a reversed-phase C18 HPLC column. Peptide-containing fractions were analyzed on a 5600+ TripleTOF (SCIEX) mass spectrometer, operating in information-dependent acquisition mode via an online Eksigent Tempo nanoLC autosampler (SCIEX) and Eksigent chPLC nanoflex (SCIEX) system. For peptide identification, data were searched by the Paragon algorithm of ProteinPilot \((4.5; \text{SCIEX})\) against the VACV strain WR proteome \((2012-01; \text{Uniprot})\). We initially searched against the VACV proteome alone, making the following choices when analyzing the
seqenced spectra. (I) Peptide FDR was ignored, instead allowing for a ProteinPilot confidence score of as low as 50% to be acceptable for subsequent testing. (II) Peptides that were detected as modified were included, but considered as unmodified. To account for potential false-positives deriving from mouse-derived pMHC sequences, all spectra were sequenced against the Mus musculus proteome (2018–08; Uniprot) appended to the above VACV proteome. Spectra matching to mouse sequences in the latter search and VACV sequences in the former search were excluded, subject to further validation.

Correlation of Synthetic Spectra with Observed Eluted Spectra. Synthetic peptides were analyzed under LC-MSMS conditions as above. MS2 spectra from each synthetic peptide were then compared with that of the original identified peptide from VACV-infected DC2.4 cells. Similarly between the spectra was assessed manually and by taking the log_{10} intensities of identified b and y ions and calculating their Pearson correlation coefficient and corresponding P-value.

MMR. A mixture of each synthesized isotopic heavy peptide was spiked into the acid eluate of immunoprecipitated pMHC complexes from VACV- or mock-infected cells before fractionation by RP-HPLC. Samples were assessed for the presence of peptide by MRM, acquired on a QTRAP 5500 (SCIEX) mass spectrometer equipped with an Eksigent TempnanoLC (SCIEX) autosampler and an Eksigent chipLC nanoﬂex (SCIEX) and utilizing Analyst 1.6 (SCIEX) software. MMR transitions are listed in Dataset S1. Data were analyzed in Skyline 64-bit [v4.1.0.1816], MacCoss Laboratory (35), with dot-product values reported in Dataset S1.

Measurement of H-2Kb-Binding Affinity. Measurements of peptide affinities for murine class I H-2Kb and Dd molecules were based on the inhibition of binding of high-affinity radiolabeled peptides to purified MHC molecules and performed as detailed elsewhere (36). Each competitor peptide was tested at six concentrations over a 10^{-1}-fold range and in three or more independent experiments. Measured IC_{50} values using this method are reasonable approximations of true K_{b} (37, 38).

Quantification of Peptide-Specific CD8+ T Cells. Splenocytes from mice infected with VACV 7 d previously were used to determine CD8+ T cell responses to peptides (23–25). Briefly, 1 × 10^{6} splenocytes were stimulated with peptides at 10^{-4} M for a total of 4 h (with 5 μg/mL Brefeldin A for the last 3 h), before being stained with monoclonal antibodies for surface CD8 and intracellular IFN-γ. Negative controls had no peptide added. The average plus three SDs (of IFN-γ as a percent of CD8+) of negative controls was the threshold to determine whether a particular peptide was immunogenic in a test. To estimate the size of the response to each peptide, the average from the background wells was subtracted from values for each well. To determine the total size of anti-VACC CD8+ T cell responses, splenocytes were stained directly ex vivo for surface CD8 and CD62L and intracellular granule B.

Data Analysis and Statistics. As described above, mass spectra were analyzed using ProteinPilot v4.5 (SCIEX). Raw mass spectra were analyzed using Peakview v2.2 (SCIEX). MHC-binding predictions (IC_{50}) were done using IDEB analysis resources NetMHConv (ver. 1.1) and NetMHCCo4.0 (39). Statistical analyses of data were carried out in GraphPad Prism v7.01 or R studio using the tests noted in the text and/or figure legends. For statistical tests, P < 0.05 was considered to be significant; unless noted otherwise, error bars or graphs denote SD.

ACKNOWLEDGMENTS. We thank the staff of the ANU Australian Phenomics Facility for animal husbandry. This work was supported by a Project Grant from the National Health and Medical Research Council Australia (NHMRC) (App1103008) (to N.L.L.G.); an NHMRC Senior Research Fellowship (AP1104329) (to D.C.T.); an NHMRC Principal Research Fellowship (AP11137739) (to A.W.P.); and a Viertel Fellowship, ARC Future Fellowship, and NHMRC Program Grant (APP1071918) (to N.L.L.G.).
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Croft, NP; Smith, SA; Pickering, J; Sidney, J; Peters, B; Faridi, P; Witney, MJ; Sebastian, P; Flesch, IEA; Heading, SL; Sette, A; La Gruta, NL; Purcell, AW; Tscharke, DC

Title:
Most viral peptides displayed by class I MHC on infected cells are immunogenic

Date:
2019-02-19

Citation:
Croft, N. P., Smith, S. A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Witney, M. J., Sebastian, P., Flesch, I. E. A., Heading, S. L., Sette, A., La Gruta, N. L., Purcell, A. W. & Tscharke, D. C. (2019). Most viral peptides displayed by class I MHC on infected cells are immunogenic. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 116 (8), pp.3112-3117. https://doi.org/10.1073/pnas.1815239116.

Persistent Link:
http://hdl.handle.net/11343/250417

File Description:
published version

License:
CC BY-NC-ND