

ANTIGEN PRESENTATION

Comment on “A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands”

Zach Rolfs^{1*}, Markus Müller^{2*}, Michael R. Shortreed¹, Lloyd M. Smith^{1†}, Michal Bassani-Sternberg^{3,4†}

There is still no convincing evidence for the frequent occurrence of posttranslationally spliced HLA-I peptides.

Mass spectrometry–based immunopeptidomics is the method of choice for characterizing peptides bound to human leukocyte antigen (HLA). Recently, Faridi *et al.* (1) reported that a large fraction of HLA-I ligands, between 13 and 45%, are derived from posttranslational cis- and trans-splicing, consistent with a recent *Science* publication from a different group reporting that 30% of the HLA-I ligandome originates from proteasomal cis-splicing (2). However, only a few such putative spliced peptides have been validated through extensive molecular and biochemical validations (3).

Faridi *et al.* used the PEAKS (4) *de novo* sequencing software to identify a large fraction of the HLA-I ligandome as cis- or trans-spliced peptides. The results are in contrast to our recent publications where we also applied *de novo* and dedicated spliced peptide identification approaches and estimated that no more than 1 to 6% of the peptides could be cis-spliced (5, 6) and that the contribution of trans-spliced peptides is likely to be much lower. We sought to evaluate this discrepancy by reanalysis of the Faridi *et al.* data. We selected three representative alleles (HLA-A*01:01, HLA-A*02:03, and HLA-B*07:02) from the 17 studied by Faridi *et al.* and analyzed all the available data for each.

We examined the published data and results using three different approaches: (i) comparison of predicted and observed hydrophobicity for the reported spliced and nonspliced peptides, (ii) reanalysis of the data using the same PEAKS Studio 8.5 *de novo* software used originally, and (iii) reanalysis of the data using the widely used search engine Comet (7). We conclude from this analysis that the fraction of spliced peptides reported was vastly overestimated.

HYDROPHOBICITY ANALYSIS

Peptide hydrophobicity is an established and orthogonal parameter for validation of peptide identifications. Faridi *et al.* applied reverse-phase liquid chromatography to separate peptides based on their hydrophobicities before tandem mass spectrometry (MS/MS). Peptide hydrophobicity is readily calculated using widely available software, and the results may be compared with chromatographic retention times for validation. We used SSRCalc (8) (version Q) to predict hydrophobicities for all HLA-A*01:01 peptides reported by Faridi *et al.* in “A0101.pep.xml” from “721-221-A1 A6810 September 13 2011 Nathan 721-221-A1.wiff” and “721-221-A1 A579 September 13 2011

Nathan 721-221-A1.wiff.” The predicted hydrophobicities for the nonspliced peptides exhibit the expected tight correlation with their observed retention times. In contrast, the predicted hydrophobicities and retention times for spliced peptides are not correlated, indicative of incorrect identification (Fig. 1A). We propose on the basis of this simple analysis alone that the vast majority, if not all, of the spliced peptide identifications reported by Faridi *et al.* are likely incorrect.

PEAKS REANALYSIS

PEAKS provides a score for each peptide, which is a measure of confidence for each identification. One expects the score distributions of spliced and nonspliced peptides to match for a given confidence if both sets are well behaved. We extracted all peptides reported by Faridi *et al.* for the HLA-A*01:01 sample and plotted the PEAKS score distributions for the count of spliced and nonspliced peptides with each score (Fig. 1B, left y axis), as well as the ratio of spliced to total peptides at or above each score (Fig. 1B, right y axis). The observed strong decrease in spliced peptide identifications compared with nonspliced identifications at high scores shows that the sets do not behave similarly and indicates a preponderance of incorrect identifications in the spliced set.

The approach used by Faridi *et al.*, wherein they added peptides identified by *de novo* sequencing to a peptide database used for a subsequent classical search, is inherently biased toward reidentification of the *de novo* identified peptides and results in a commensurate underestimation of the peptide false discovery rate (FDR). Confidence thresholds can be readily obtained for database searches, but confidence estimates are more difficult for *de novo* sequencing, and the FDR is frequently high (9). We provide here an alternative means to estimate FDR that still makes use of the PEAKS database search used in the Faridi *et al.* publication. We performed a PEAKS database search using only nonspliced, UniProt-curated peptides in the database to obtain a set of high-confidence, gold standard peptide identifications. We then used this set of gold standard peptides to estimate the FDR at each PEAKS *de novo* score as previously described (10). The ratio of reported spliced peptides to total peptides is plotted in Fig. 1C as a function of the estimated FDR. The plot shows a negligible number of spliced identifications at high confidence (i.e., at low FDR). Faridi *et al.* reported for HLA-A*02:03 that 23.5% of the peptides detected were spliced at about 1% FDR; however, we find that less than 1% of the reported peptides could potentially be spliced at a 1% FDR.

ALTERNATE SEARCH ENGINE ANALYSIS

In a final attempt to reidentify spliced peptides in the data reported by Faridi *et al.*, we analyzed the data using the well-established and

¹Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA. ²Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. ³Ludwig Cancer Research Center, University of Lausanne, 1066 Epalinges, Switzerland. ⁴Department of Oncology, University Hospital of Lausanne, 1011 Lausanne, Switzerland.

*These authors contributed equally to this work.

†Corresponding author. Email: michal.bassani@chuv.ch (M.B.-S.); smith@chem.wisc.edu (L.M.S.)

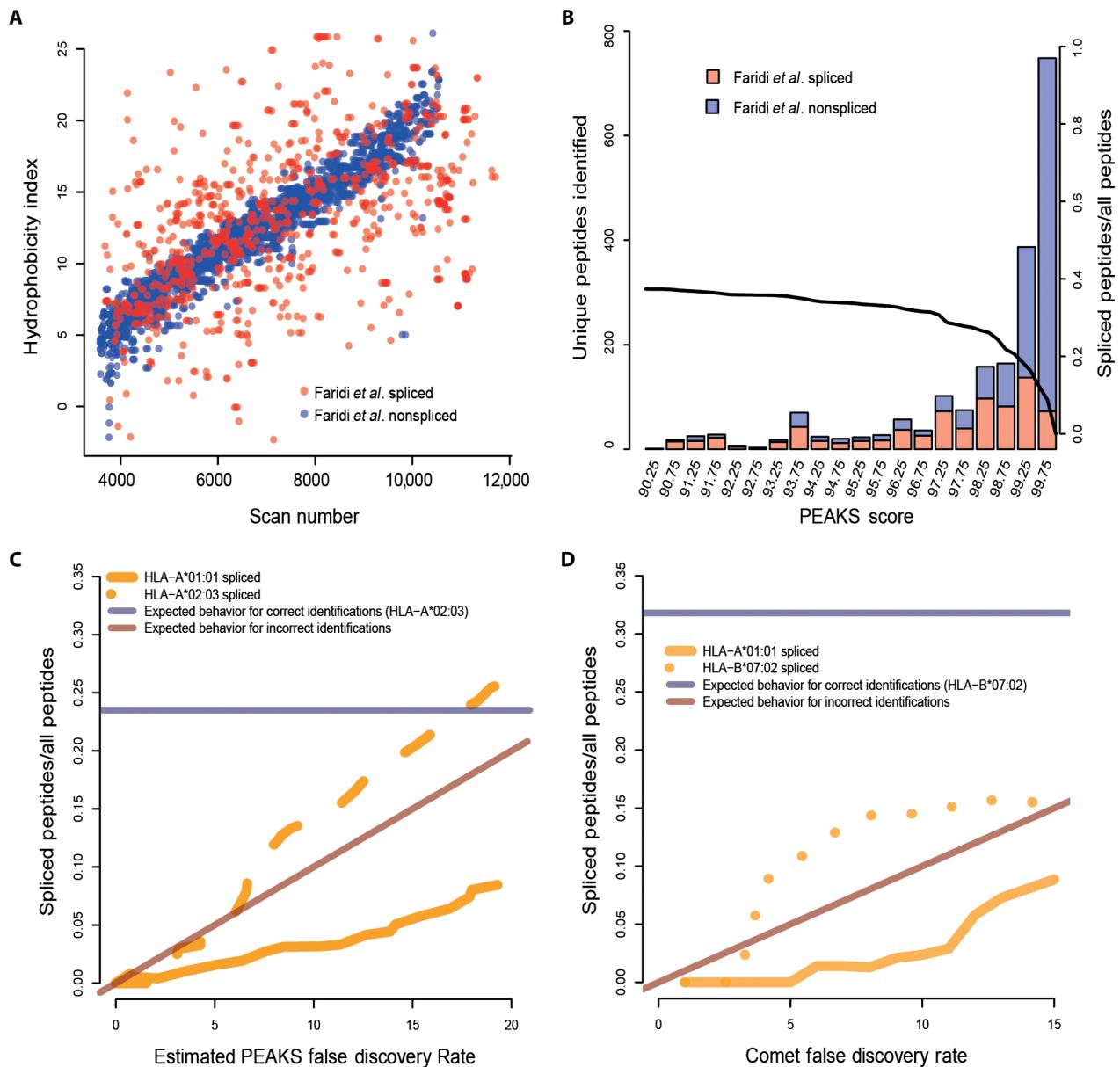


Fig. 1. Analysis of a subset of the data from Faridi *et al.* shows no evidence for a substantial number of posttranslationally spliced HLA-I peptides. (A) Predicted hydrophobicity index against the MS/MS scan numbers for nonspliced (blue) and spliced (red) peptide sequences identified by Faridi *et al.* for the HLA-A*01:01 sample. (B) Distribution of unique nonspliced (blue) and spliced (red) peptide sequences identified by Faridi *et al.* for the HLA-A*01:01 sample as a function of the PEAKS de novo score. The ratio of spliced peptides to all identified peptides at or above each score is shown with a line. (C) Percentage of the spliced peptide contribution to the ligandome found by our reanalysis of the HLA-A*01:01 and HLA-A*02:03 samples with the PEAKS software as a function of the error level. (D) Percentage of the spliced peptide contribution to the ligandome found by our reanalysis of the HLA-A*01:01 and HLA-B*07:02 samples with the Comet software as a function of the Comet error level.

freely available search engine Comet (7). In this search, we used a custom database containing all UniProt human proteins (version 26.09.18, reviewed with isoforms) and all the reported spliced peptides for the HLA-A*01:01 or HLA-B*07:02 samples. Comet was run with the search settings reported by Faridi *et al.* The XCorr score and reversed protein decoy sequences were used for FDR calculation. At low FDR (<4%), Comet could not identify any spliced peptides. Even at FDR values as high as 15%, the percentage of spliced peptides was much lower than that reported by Faridi *et al.* (Fig. 1D).

MOTIF ANALYSIS AND SYNTHETIC PEPTIDE VALIDATIONS

Motif analysis was used by Faridi *et al.* to validate the spliced peptides, but there were notable differences between the motifs of the spliced and nonspliced peptides. Independent of the results, motif analysis can only be used to verify that the anchor residues were correctly identified, not that the entire sequence is correct. In addition, several images of synthetic peptide spectra were offered as validation, but many of these appear to differ substantially from the experimental spectra they sought to validate.

In summary, we have been unable to find evidence for a substantial number of spliced peptides in the data reported by Faridi *et al.* It is important to note that our analysis is based on a subset of the data, specifically the raw data for HLA alleles A*01:01, A*02:03, and B*07:02 and the results file for A*01:01. Of the 17 different alleles analyzed by Faridi *et al.*, only eight results files were made publicly available on PRIDE (PXD009660), and only three of those eight results files contained spliced peptides (“A0101.pep.xml,” “B0702.pep.xml,” and “B5801.pep.xml,” as of 2 May 2019). Requests to Faridi *et al.* for additional results files were unsuccessful. Spliced peptide sequences for all alleles were reported by Faridi *et al.* in table S4 (1), but the MS/MS scan numbers that these sequences were identified from were not reported. Results files are necessary to evaluate the putative spliced peptide identifications, because they contain the scan numbers and PEAKS scores needed to determine retention times (Fig. 1A) and PEAKS score distributions (Fig. 1B). Nevertheless, from our analysis of this subset of data, we believe that it is clear that the authors incorrectly interpreted their data, leading them to erroneously conclude that a high proportion of spliced peptides was present in their samples.

REFERENCES AND NOTES

1. P. Faridi, C. Li, S. H. Ramarathinam, J. P. Vivian, P. T. Illing, N. A. Mifsud, R. Ayala, J. Song, L. J. Gearing, P. J. Hertzog, N. Ternette, J. Rossjohn, N. P. Croft, A. W. Purcell, A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci. Immunol.* **3**, eaar3947 (2018).
2. J. Liepe, F. Marino, J. Sidney, A. Jeko, D. E. Bunting, A. Sette, P. M. Kloetzel, M. P. H. Stumpf, A. J. R. Heck, M. Mishto, A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **354**, 354–358 (2016).
3. N. Vigneron, V. Stroobant, V. Ferrari, J. Abi Habib, B. J. Van den Eynde, Production of spliced peptides by the proteasome. *Mol. Immunol.* **18**, 10.1016/j.molimm.2018.03.030 (2018).
4. J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, B. Ma, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).
5. R. Mylonas, I. Beer, C. Iseli, C. Chong, H.-S. Pak, D. Gfeller, G. Coukos, I. Xenarios, M. Müller, M. Bassani-Sternberg, Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol. Cell. Proteomics* **17**, 2347–2357 (2018).
6. Z. Rolfs, S. K. Solntsev, M. R. Shortreed, B. L. Frey, L. M. Smith, Global identification of post-translationally spliced peptides with neo-fusion. *J. Proteome Res.* **18**, 349–358 (2019).
7. J. K. Eng, T. A. Jahan, M. R. Hoopmann, Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
8. O. V. Krokhin, S. Ying, J. P. Cortens, D. Ghosh, V. Spicer, W. Ens, K. G. Standing, R. C. Beavis, J. A. Wilkins, Use of peptide retention time prediction for protein identification by off-line reversed-phase HPLC–MALDI MS/MS. *Anal. Chem.* **78**, 6265–6269 (2006).
9. N. H. Tran, X. Zhang, L. Xin, B. Shan, M. Li, De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8247–8252 (2017).
10. A. Devabhaktuni, J. E. Elias, Application of de Novo sequencing to large-scale complex proteomics data Sets. *J. Proteome Res.* **15**, 732–742 (2016).

Funding: This work was supported by the National Cancer Institute of the NIH under award number U24CA199347. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was supported by the Ludwig Institute for Cancer Research and by the ISREC Foundation thanks to a donation from the Biltema Foundation. **Competing interests:** The authors declare that they have no competing interests.

Submitted 5 December 2018
Accepted 16 July 2019
Published 16 August 2019
10.1126/sciimmunol.aaw1622

Citation: Z. Rolfs, M. Müller, M. R. Shortreed, L. M. Smith, M. Bassani-Sternberg, Comment on “A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands.” *Sci. Immunol.* **4**, eaaw1622 (2019).

Comment on "A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands"

Zach Rolfs, Markus Müller, Michael R. Shortreed, Lloyd M. Smith and Michal Bassani-Sternberg

Sci. Immunol. 4, eaaw1622.

DOI: 10.1126/sciimmunol.aaw1622

ARTICLE TOOLS

<http://immunology.sciencemag.org/content/4/38/eaaw1622>

REFERENCES

This article cites 9 articles, 5 of which you can access for free
<http://immunology.sciencemag.org/content/4/38/eaaw1622#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Immunology (ISSN 2470-9468) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Immunology* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works