The Choice of HLA-Associated Peptide Enrichment and Purification Strategy Affects Peptide Yields and Creates a Bias in Detected Sequence Repertoire

Annalisa Nicastri,* Hanqing Liao, Julius Muller, Anthony W. Purcell, and Nicola Ternette*

Understanding the most appropriate workflow for biochemical human leukocyte antigen (HLA)-associated peptide enrichment prior to ligand sequencing is essential to achieve optimal sensitivity in immunopeptidomics experiments. The use of different detergents for HLA solubilization as well as complementary workflows to separate HLA-bound peptides from HLA protein complex components after their immunoprecipitation including HPLC, C18 cartridge, and 5 kDa filter are described. It is observed that all solubilization approaches tested led to similar peptide ligand identification rates; however, a higher number of peptides are identified in samples lysed with CHAPS compared with other methods. The HPLC method is superior in terms of HLA-I peptide recovery compared with 5 kDa filter and C18 cartridge peptide purification methods. Most importantly, it is observed that both the choice of detergent and peptide purification strategy creates a significant bias for the identified peptide sequences, and that allele-specific peptide repertoires are affected depending on the workflow of choice. The results highlight the importance of employing a suitable strategy for HLA peptide enrichment and that the obtained peptide repertoires do not necessarily reflect the true distributions of peptide sequences in the sample.

Peptides presented on the human leukocyte antigen (HLA) class I molecules (HLAp) on the cellular surface represent the status of the intracellular environment and any perturbation due to pathogenic infection or disease. HLAp are continuously scrutinized by CD8+ cells, and the identification of new T-cell epitopes is a crucial step for the development of novel therapies and vaccines against infection, cancer, metabolic and autoimmune disorders. Over the last three decades, many research groups have established different methods to extract HLAp for subsequent analysis and sequence identification using liquid chromatography mass spectrometry (LC-MS). Two main approaches have been described: mild acid elution (MAE)[1] and immunoprecipitation (IP)[2] of soluble or membrane-bound HLA.[3,4] MAE is infrequently used due to the high identification of non-specific peptides (40–60%) and due to the inapplicability of the methodology to tissues and frozen samples.[5,6] IP with allo- and serotype specific HLA antibodies is therefore the most common and more specific method employed for the extraction of HLA-bound peptides.

Here, we report a comparison of the use of different detergents for HLA solubilization as differences were observed by others,[7] in addition to complementary approaches to separate HLA-bound peptides from HLA protein complex components after immunoprecipitation.

We first compared the use of four different detergents for HLA solubilization. We evaluated the use of the most commonly used non-ionic detergent IGEPAL CA-630 (Igepal) and the closely related Triton X-100 (Triton), the zwitterionic surfactant 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), and the ionic detergent sodium deoxycholate (DOC). We further included a control IP in which no detergent was used (Ctrl). We carried out the experiments using either half or twice the reported critical micelle concentration (CMC) for each detergent (further referred to as low-CMC and high-CMC samples). For the DOC, we used a single published concentration of 0.25%[8] (Table 1). The experiment was conducted in two biological replicates (with exception of CHAPS at low concentration, for which one sample was lost) and two technical replicate acquisitions by LC-MS. In brief, 3 × 10⁸ Jurkat cells for each sample were lysed in 10 mL 150 mM NaCl, 50 mM Tris-HCl pH 8 supplemented with a Protease Inhibitor Cocktail (EDTA-free, Roche) using the...
different detergents described previously (Figure S1, Supporting Information and Table 1). The HLA-I molecules were purified by immunoprecipitation (IP) using anti-pan HLA-I monoclonal antibody (clone W6/32) and eluted with 10% acetic acid. The HLAp were then separated from the HLA-I molecules by HPLC and analyzed with an Ultimate 3000 UPLC coupled to an Orbitrap Fusion Lumos instrument (Thermo Scientific). Detailed information on the purification protocol and the acquisition parameters are described in Purcell et al. Finally, the peptide sequences were identified and quantified by PEAKS 8.5 software (Bioinformatic Solutions Inc.), at a chosen score cut-off of 10 logP (Bioinformatic Solutions Inc.), at a chosen score cut-off of 0.74% resulting in the most identifications (4420 peptide sequences identified on average), followed by Igepal 0.1% (3617 peptide sequences identified on average), Triton 0.25% (38.19%) and DOC 0.25% (38.04%). These differences were confirmed in the overall assessment of binding rank score distributions (Figure 1G).

Most peptides were identified in the high-CMC samples, with CHAPS 0.74% resulting in the most identifications (4420 peptide sequences identified on average), followed by Igepal 0.1% (4205 peptide sequences identified on average), Triton 0.1% (3750 peptide sequences identified on average), and DOC 0.25% (3617 peptide sequences identified on average) (Figure 1A). In comparison, we identified significantly less peptides from the low-CMC samples, and in the ctrl samples that were processed without the addition of detergent. We also found that the high-CMC samples were associated to a higher overall abundance of HLAp signal intensity compared with samples in which lower concentrations of detergent were used (Figure 1B), demonstrating higher HLAp yields. High-CMC samples showed a 52–62% overlap in peptide sequences between technical replicates, and 45–64% between biological replicates (Figure S2, Supporting Information). All samples showed similar length distributions with >50% peptides being 9-mers, whereas we observed a wider distribution of longer peptides in the control (Figure 1C). Overall, 83.1% of peptides of a length between 8 and 14 amino acids were predicted to bind to at least one of the five common HLA-I alleles. For this, we assigned the allele with the minimal rank score as defined by NetMHCpan 4.0, and compared the resulting peptide numbers between the different samples (Figure 1F and Table S1, Supporting Information). No apparent differences were detected; however, HLAp uniquely detected in individual high-CMC samples showed some minor variations: a larger repertoire of sequences predicted to bind to B*07:02 and B*35:03 were identified using CHAPS 0.74% (22.08% and 14.6%, respectively), compared with the other detergents, and peptides assigned to HLA-A*03:01 were preferentially extracted when Triton 0.1% was used (38.19%). A higher frequency of predicted non-binding peptides were detected in samples processed with DOC 0.25% (47.53%) and Igepal 0.1% (38.04%). These differences were confirmed in the overall assessment of binding rank score distributions (Figure 1G).

To decipher whether the sequence bias causing the shift in allele assignment was significant, we assessed individual amino acid frequency distributions at the two HLA anchor positions, the second and the C-terminal amino acid for all 8–14mer sequences identified in high-CMC samples (Figure 1H).

Interestingly, for the binding anchor at position 2, we noticed a significantly higher frequency of proline (P) in the peptides identified in the sample CHAPS 0.74% compared with Igepal 0.1% and DOC 0.25%. Higher frequency of P was also identified in the Triton 0.1% sample and lower frequency in both DOC 0.25% and Igepal 0.1%, although these results were not significant. This increased frequency explains the higher proportion of peptides assigned to both HLA-B*07:02 and B*35:03 in CHAPS 0.74% as described earlier, as these alleles require a P as anchor in position 2.

Finally, we assessed the proportion of histone peptides that are commonly observed contaminants in all samples. We found a higher quantity of histone peptides in the DOC samples compared with the other detergents (Figure 1I). We then proceeded to evaluate the second arm of the IP protocol and compared three different strategies to purify HLAp from the larger complex components. We applied the commonly used HPLC purification (Thermo Scientific Ultimate 3000 supplemented with a ProSwift RP-1S 4.6 × 50 mm column), C18 cartridge purification (SepPak, Waters), and 5 kDa filter (Millipore) separation. Starting from 1 × 10⁷ Jurkat cells for each biological replicate, samples were processed as described.

### Table 1. Overview of detergents and critical micelle concentration (CMC).

| Detergent | Class       | M_r [g mol⁻¹] | Conc. [% [w/v]] | Conc. [mM] | CMC [mM] |
|-----------|-------------|---------------|-----------------|------------|----------|
| Igepal    | Non-ionic   | Varying       | 0.025           | 0.044      | 0.083–0.088 |
|           |             |               | 0.1             | 0.176      |          |
| Triton X-100 | Non-ionic | 625           | 0.006           | 0.1        | 0.2–0.9 |
|           |             |               | 0.113           | 1.8        |          |
| CHAPS     | Zwitterionic | 615           | 0.74            | 12         | 6–10     |
|           |             |               | 0.18            | 3          |          |
| DOC       | Ionic       | 415           | 0.25            | 6.0        | 2–6      |

[4] https://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/solubilisation_membrane_proteins/
Figure 1. Comparison of the use of different detergents in HLAp enrichment. A) Identified peptide sequences in each biological replicate, error bars are technical variation from repeated LC-MS acquisition. B) Peptide log₁₀ intensity distribution; boxes represent the interquartile range of the median; C) Peptide length distribution; D) Sequence motifs for 9mer peptides generated from combined datasets of all samples as defined by NetMHCpan4.0. E) Sequence overlap in high-CMC samples. F) HLA binding prediction of peptides identified in each condition using NetMHCpan, and uniquely identified in each condition (indicated by U_). The allele with the lowest rank score and therefore highest likelihood of binding for each peptide sequence is plotted. G) Rank score (log scale) distributions reflecting the likelihood of binding for all (All) and uniquely (Unique) identified peptides in each condition. Rank scores below 2 indicate binding as indicated by the red line. H) Overall amino acid frequency distribution across all 8–14mers in individual samples (All), all uniquely identified peptides across conditions (Unique), position 2 of the peptide sequence (P2), and the C-terminal position (PC). I) Intensity distribution of peptide derived from histones across all samples.
Figure 2. Comparison of the use of different peptide purification strategies. A) Identified peptide sequences in each biological replicate, error bars are technical variation from repeated LC-MS acquisition. B) Peptide log10 intensity distribution; boxes represent the interquartile range of the median; C) Peptide length distribution; D) Sequence motifs for 9mer peptides generated from combined datasets of all samples as defined by NetMHCpan4.0. E) Sequence overlap in high-CMC samples. F) HLA binding prediction of peptides identified in each condition using NetMHCpan, and uniquely identified in each condition (indicated by U). The allele with the lowest rank score and therefore highest likelihood of binding for each peptide sequence is plotted. G) Rank score (log scale) distributions reflecting the likelihood of binding for all (All) and uniquely (Unique) identified peptides in each condition. Rank scores below 2 indicate binding as indicated by the red line. H) Overall amino acid frequency distribution across all 8–14mers in individual samples (All), all uniquely identified peptides across conditions (Unique), position 1 of the peptide sequence (P1), position 2 of the peptide sequence (P2) and the C-terminal position (PC). I) Average isoelectric point of all peptides and uniquely (U) identified peptides per condition.
previously. After IP and acid elution, the sample was divided into three equal amounts, which were purified by HPLC (10 min gradient 3–28% buffer B (0.1% TFA in acetonitrile)) in buffer A (0.1% TFA in water), C18 cartridge (peptides were eluted with 28% ACN, 0.1% TFA in water), and 5 kDa filter, respectively (Figure S3, Supporting Information). Experiments were performed in two technical and two biological replicates and analyzed as stated previously. The FDR range was 0.8–2.0% (average 1.4%) at a score cut-off of $-10 \log P = 15$.

HPLC purification allowed the identification of significantly more peptides compared with C18 and 5 kDa filter (40.7% and 28.9%, respectively), and more peptides were identified with the 5 kDa filter method compared with C18 purification (Figure 2A). Peptide abundances were higher overall in samples in which HPLC was used compared with HLAp obtained after C18 and 5 kDa filter purification (Figure 2B), and we observed a higher frequency of 9-mer peptides for the HPLC purified samples (Figure 2C). Samples showed a 57–71% overlap in peptide sequences between technical replicates, and 54–64% for the biological replicates (Figure S4, Supporting Information). Overall, 88.5% of 8–14mer peptides were predicted to bind to at least one of the five classical Jurkat alleles (Figure 2D), and respective sequence motifs represented the expected amino acid distribution for all alleles. Only 30.8% of HLAp were commonly identified across the three methods, whereas 21%, 7.2%, and 12.6% represented unique peptides for HPLC, C18, and 5 kDa filter, respectively (Figure 2E).

We could again observe biases in identified HLA-I reports that were subsequently assigned to a given HLA I alloype, when comparing overall and uniquely identified HLAp sequences across the analyzed conditions (Figure 2F and Table S2, Supporting Information). Fewer HLA-A*03:01 peptides were identified in samples C18 in comparison to application of both HPLC and 5 kDa purification methods. Higher numbers of HLA-B*07:02, B*35:03, and C*04:01-assigned peptides were found uniquely to the HPLC purified samples (19.2%, versus 12.6% in the 5 kDa sample and 5.24% in the C18 sample, respectively). Peptides assigned to HLA-C*07:02 were better represented in the sample enriched using the C18 cartridges (8%). Overall rank score distributions showed generally lower binding scores for peptides uniquely identified in the sample enriched with the C18 cartridges, suggesting this strategy was the lowest performing in our study (Figure 2G).

In order to further understand if the different separation strategies impact upon the nature of the peptides available for LC-MS/MS analysis, we examined changes in the distribution of amino acids at different positions in the identified peptides. Interestingly, peptides detected uniquely in the 5 kDa filter sample were characterized by a significantly higher frequency of basic residues R and K at position 1 in comparison with C18 cartridge and HPLC enrichment. Peptides uniquely identified in the HPLC sample were characterized by significantly higher occurrence of P in position 2 (Figure 2H). Peptides with R at P2 were detected with higher frequencies in C18 cartridge purified samples, and R was also detected with higher frequency at the C-terminus in the C18 cartridge purified samples. As a result of significantly differing frequencies of detection of basic amino acids of peptides enriched using the three different approaches, the average isoelectric point (pI) values differed with highest values found in the 5 kDa sample (pI = 9.3), followed by HPLC (pI = 8.7) and C18 cartridge (pI = 8.4) (Figure 2I). One factor for this observation could be that reversed phase chromatographic material absorbs peptides with basic amino acids at the N-terminus, which could be resulting in lower yields and detection of such peptides when using either HPLC or C18 cartridge purification strategies. Alternatively, basic hydrophilic peptides could be less likely absorbed by the 5 kDa filter material. Further, we observed that peptides with P in position 2 were detected with higher frequency by HPLC.

We compared different HLAp extraction and purification strategies with the aim of optimizing the yield of HLAp in immunopeptidomics workflows. We found that CHAPS lysis at a concentration of 0.76% performed best and allowed identification of a higher number of peptides than with all other detergents tested. We would like to note that these observations may be dependent on the choice of buffer volume and resulting final protein concentration in the lysate. Furthermore, the use of HPLC for separating peptides from larger complex components was superior compared to other methods. The different purification strategies had a clear effect on yielded peptide properties. We detected significant differences in amino acid composition and binding rank score distributions for unique peptide sequences identified underlining the fact that the use of different workflows may directly affect complex stability and peptide recovery.

In conclusion, our experiments document differences and biases obtained in immunopeptidomics experiments, by direct comparison of different peptide enrichment strategies. We would like to emphasize that the presented data are the results obtained in our laboratory, and that these may vary in other laboratories where altered experimental conditions are used. We would like to highlight that the availability of a global HLA standard and a strategy to benchmark HLA peptide enrichment across laboratories is therefore essential and should be considered an important aim for the field.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
Mass spectrometry data were acquired in the Target Discovery Institute MS Laboratory led by Benedikt M. Kessler.

Conflict of Interest
The authors declare no conflict of interest.

Keywords
detergents, human leukocyte antigen class I molecules, immunopeptidomics, purification strategy

Received: December 8, 2019
Revised: April 14, 2020
Published online: June 2, 2020
[1] W. J. Storkus, H. J. Zeh 3rd, R. D. Salter, M. T. Lotze, J. Immunother. Emphasis Tumor Immunol. 1993, 14, 94.
[2] K. Prilliman, M. Lindsey, Y. Zuo, K. W. Jackson, Y. Zhang, W. Hildebrand, Immunogenetics 1997, 45, 379.
[3] D. F. Hunt, R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. L. Cox, E. Appella, V. H. Engelhard, Science 1992, 255, 1261.
[4] K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, H. G. Rammensee, Nature 1991, 351, 290.
[5] E. Caron, D. J. Kowalewski, C. Chiek Koh, T. Sturm, H. Schuster, R. Aebersold, Mol. Cell. Proteomics 2015, 14, 3105.
[6] J. Lanoix, C. Durette, M. Courcelles, E. Cossette, S. Comtois-Marotte, M. P. Hardy, C. Cote, C. Perreault, P. Thibault, Proteomics 2018, 18, 1700251.
[7] T. A. Springer, D. L. Mann, A. L. DeFranco, J. L. Strominger, J. Biol. Chem. 1977, 252, 4682.
[8] M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, M. Mann, Mol. Cell. Proteomics 2015, 14, 658.
[9] A. W. Purcell, S. H. Ramarathinam, N. Ternette, Nat. Protoc. 2019, 14, 1687.
[10] M. C. Thomsen, M. Nielsen, Nucleic Acids Res. 2012, 40, W281.
[11] M. Andreatta, M. Nielsen, Bioinformatics 2016, 32, 511.