Published in final edited form as:

Proteomics. 2017 October ; 17(19): . doi:10.1002/pmic.201700177.

Data independent acquisition of HLA class I peptidomes on the Q Exactive mass spectrometer platform

Danilo Ritz^{#1}, Jonny Kinzi^{#2}, Dario Neri^{2,\$}, and Tim Fugmann^{1,\$}

¹Philochem AG, Libernstrasse 3, CH-8112 Otelfingen, Switzerland ²Institute of Pharmaceutical Sciences, ETH Zurich, 8093 Zurich, Switzerland

[#] These authors contributed equally to this work.

Abstract

The characterization of peptides presented by human leukocyte antigen (HLA) class I molecules is crucial for understanding immune processes, biomarker discovery and the development of novel immunotherapies or vaccines. Mass spectrometry allows the direct identification of thousands of HLA-bound peptides from cell lines, blood or tissue. In recent years, data-independent acquisition (DIA) mass spectrometry methods have evolved, promising to increase reproducibility and sensitivity over classical data-dependent acquisition (DDA) workflows. Here we describe a DIA setup on the Q Exactive mass spectrometer, optimized regarding the unique properties of HLA class I peptides. The methodology enables sensitive and highly reproducible characterization of HLA peptidomes from individual cell lines.

From up to 16 DDA analyses of 100 million human cells, more than 10'000 peptides could be confidently identified, serving as basis for the generation of spectral libraries. This knowledge enabled the subsequent interrogation of DIA data, leading to the identification of peptide sets with >90% overlap between replicate samples, a prerequisite for the comparative study of closely-related specimens. Furthermore, >3'000 peptides could be identified from just 1 million cells after DIA analysis using a library generated from 300 million cells. The reduction in sample quantity and the high reproducibility of DIA-based HLA peptidome analysis should facilitate personalized medicine applications.

Keywords

DIA; HLA peptidomics; immunopeptidomics; antigens; peptidomics; biomarker

---

Human leukocyte antigen (HLA) class I complexes present peptides on the surface of all nucleated cells of the human body. Displayed peptides originate from proteasomal

---

^{\$}Corresponding Authors: Prof. Dario Neri, Institute of Pharmaceutical Sciences, HCI G 392.4, Vladimir-Prelog-Weg 1-5/10, ETH Zurich, 8093 Zurich, Switzerland, neri@pharma.ethz.ch, Phone: +41 44 633 74 01, Fax: +41 44 633 13 58; Dr. Tim Fugmann, Philochem AG, Libernstrasse 3, CH-8112 Otelfingen, Switzerland, tim.fugmann@philochem.ch, Phone: +41 43 544 88 00, Fax: +41 43 544 88 09.

Conflicts of Interest: Dario Neri is co-founder of Philogen, shareholder and member of the board. Tim Fugmann and Danilo Ritz are employees of Philochem AG. The authors declare no additional conflict of interest.

degradation products and have a predominant length between eight and twelve amino acids [1]. Immune cells can interact with HLA-peptide complexes, thereby detecting pathological events such as malignant transformation or infection with pathogens. Knowledge of the immunopeptidome may enable the development of novel therapeutic strategies such as next-generation immunotherapies or vaccines against cancer, infectious diseases and autoimmunity [2]. Currently, mass spectrometry remains the only methodology to identify thousands of HLA peptides presented *in vivo* [3]. Mass spectrometry-based characterization of HLA-bound peptides was pioneered by Rammensee and Hunt in the 1990s, while over 10'000 peptides can be sequenced with current procedures [4–6]. Until recently, the majority of HLA peptidome analyses were performed using well-established data-dependent acquisition (DDA), which is well suited for discovery proteomics, but suffers from limited analytical reproducibility inherent to its fundamental design [7] and is therefore less suited for quantification of analytes between multiple conditions [3, 8]. To overcome these limitations, data independent acquisition (DIA) workflows have been proposed for the analysis of HLA peptidomes [8, 9]. Both reports apply variations of the SWATH-MS methodology, an implementation of the DIA workflow on the TripleTOF mass spectrometer. DIA was initially described by Venabe and colleagues and has since been implemented in many laboratories [7, 10, 11]. For DIA, spectra are acquired following a predefined scheme, which is independent of abundance and distribution of analytes. The fragmentation and acquisition of all detectable ions within a predefined mass range yields a comprehensive data set for a given sample, which can be interrogated for presence of analytes with known fragmentation pattern [8]. Caron and colleagues demonstrated that the analysis of HLA peptidomes with the SWATH-MS approach clearly outperformed the DDA approach regarding reproducibility and sensitivity across several technical replicates [8].

Encouraged by these promising results, we investigated a DIA workflow for the analysis of HLA class I peptidomes on the Q Exactive mass spectrometer. While methods are not directly translatable between the TripleTOF and the Q Exactive platforms, the latter has been found to perform well for the analysis of tryptic digests [7]. A schematic representation of the general workflow for the DIA analysis of HLA class I peptidomes is shown in Figure 1A. The human mantle cell lymphoma cell line Maver-1 and the human embryonal kidney cell line HEK293 were used to test the workflow [12, 13]. Both cell lines were HLA typed by sequence-specific oligonucleotide (SSO) and sequence-specific primer (SSP) technologies resulting in an unambiguous annotation of the A, B and C alleles as follows: Maver-1: HLA-A*24:02, HLA-A*26:01, HLA-B*38:01, HLA-B*44:02, HLA-C*05:01 and HLA-C*12:03; HEK293: HLA-A*03:01, HLA-B*07:02 and HLA-C*07:01. In agreement with previous literature, the HEK293 HLA typing shows homozygosity regarding the A, B and C alleles [14]. HLA class I peptides were isolated from Maver-1 and HEK293 cell lysates [15]. HLA complexes were purified from cell lysates with the pan-HLA class I specific antibody W6/32. Peptides were eluted under acidic conditions and purified with a single C18 step. iRT peptides (Biognosys, Schlieren) were added before analysis on a Q Exactive mass spectrometer coupled to an Easy-nLC 1000 equipped with a 15 cm Acclaim PepMap RSLC C18. Peptides were eluted over a linear gradient from 0% to 30% ACN over 120 minutes. Spectral libraries were generated from SEQUEST search results filtered to 1% FDR using SpectraST. Transition lists created with spectrast2tsv were imported into Skyline

for analysis of DIA data. Peaks were reintegrated using mProphet and filtered to a q-value of 0.01. A detailed description of the workflow can be found in the Supporting Material and Methods.

As a starting point for our investigations, we evaluated a DIA methodology described by Caron et al. [8], who observed >95% of peptide identifications between 400 and 700 m/z [Figure 1B, left]. However, since the Q Exactive platform allows the confident identification of many peptides within an extended m/z range [Figure 1C], we compared an adaptation of Caron's method to one featuring a precursor mass range from 250 to 850 m/z [Figure 1B, center and right]. Both methods had 20 DIA windows, as this procedure was found to be optimal in terms of data points per peak. As expected, we identified many peptides below 400 m/z, without experiencing negative effects due to the larger DIA window size [Figure 1D]. Inspired by Bruderer et al., [7] we also tested a setup with variable windows between 250 and 850 m/z, with narrow windows in the m/z range in which most peptides were identified [Figure 1B, right]. However, while the variable method indeed led to the identification of more peptides in the m/z range in which the windows were smaller than 30 Da, fewer peptides were identified beyond 600 m/z, the m/z range in which the windows were larger than 40 Da [Figure 1E]. For the analysis of the HLA peptidome of Maver-1 cells, the method with fixed windows was superior to the variable method. Interestingly, the opposite was true for HEK293 cells where the variable method led to 51 additional identifications [Figure 3A]. One possible explanation for this observation is the HLA A*03:01 allele expressed by HEK293 cells, an allele predominantly binding peptides with a C-terminal arginine or lysine residue. As a result, HEK293 peptides feature higher charge states and lower m/z ratio than Maver-1 peptides [compare Figure 3B with Figure 1C]. Thus, while for general applications we prefer to use DIA implementations with fixed windows, the analysis of certain cell lines may benefit from a window size optimization procedure.

After having established an optimized DIA method for HLA peptidome analysis, we compared DDA and DIA results in terms of reproducibility and sensitivity. In total 5686, 5465, and 5528 peptides were identified from a triplicate DDA analysis of the Maver-1 HLA peptidome, with 3761 peptides (50.61%) found in all three samples [Figure 2A, left]. The corresponding analysis applying the DIA method with fixed windows resulted in the identification of 7509, 7463, and 7512 peptide sequences, with a total of 7307 peptides (95.53%) found in all three replicates [Figure 2A, right and Supporting Information Table 1A]. The comparison of the peptide intensity (i.e. Skyline total area) found in two replicates demonstrated an excellent correlation ($R^2=0.948$) [Figure 2B]. Furthermore, we observed that the intensities of identified peptides spanned five orders of magnitude, demonstrating the sensitivity of the DIA methodology for HLA peptidomics on the Q Exactive [Figure 2C]. Peptide binding prediction by NetMHCpan 3.0 is in agreement with the HLA typing of the Maver-1 cell line, indicating the quality of the data [Figure 2D]. Peptides predicted to bind to the A and B alleles were predominant, with peptides predicted to bind to C alleles less abundant, possibly due to the higher degeneracy and lower expression of C alleles [16]. Starting from these promising results, we investigated on the sensitivity of the methodology, purifying HLA class I complexes from lysate equivalent to 100, 25, 10, 5, and 1 million cells, respectively [Figure 2E]. Surprisingly, the cumulative number of peptide identifications remained essentially stable decreasing the input to 10 million cells (6401

peptides, 93.9% of the sequences identified from 100 million cells), while 85.5 and 50.3% of peptide sequences were recovered from just 5 and 1 million cells, respectively. Finally, we investigated whether a higher number of peptide sequences could be identified, by creating larger spectral libraries from additional Maver-1 samples [Supporting Information Figure 1A]. Indeed, the use of combined spectral libraries from six, twelve or sixteen samples led to a maximum of 10901 peptide identifications [Figure 2F]. Importantly, more than 90% of peptide sequences were shared between all three replicates, leading to an increase in the number of shared peptides by 8, 31, and 38%, with DIA libraries obtained from six, twelve and sixteen samples, respectively [Figure 2G]. Furthermore, the resulting peptidomes were comparable in terms of peptide length distribution and HLA-binding prediction [Supporting Information Figure 1C and D]. Similar results were obtained for HEK293 cells, for which we created spectral libraries from up to 24 HEK293 samples, allowing the identification of 7313 peptide sequences from 100 million cells. [Figure 3C-G, Supporting Information Table 1B and Supporting Information Figure 1B, E and F].

In summary, we established a DIA workflow for the analysis of HLA peptidomes, on the broadly available Q Exactive platform. Using an optimized DIA method, highly reproducible HLA peptidomes were generated, with peptide identifications spanning 5 orders of magnitude, in terms of MS signal intensity. Due to the increased sensitivity of the DIA methodology, the use of 1 million cells yielded approximately 50% of the peptide identifications obtained with 100 million cells. Additionally, an increase in spectral library size allowed the identification of even higher numbers of HLA-bound peptides, without affecting the quality of the results. At the same time, one of the main limitations of the DIA workflow clearly relates to the requirement for large, high quality spectral libraries. Currently, spectral libraries may be populated by performing database searches with DIA-Umpire [17] or through community efforts such as the SWATH atlas [<http://www.swathatlas.org>], as an alternative to the labour-intensive performance of a large number of DDA runs from cells, tissue and disease states. However, applicability of approaches such as DIA-Umpire was not demonstrated for HLA peptidomes and we found the retention time of public spectral libraries to be variable from the one observed in our experimental setup. Until these issues are solved, it will be required to create large spectral libraries for specific experimental goals with extensive efforts.

Nevertheless, the high reproducibility and sensitivity of the DIA methodology should not only prove useful to address more basic questions e.g. regarding the variability of the HLA peptidome in health and disease, but should also allow to answer why the immune system of some patients can reject established tumors after checkpoint blockade. Further, we believe that the HLA peptidome will facilitate the discovery of biomarkers for patient stratification, for monitoring response to immunotherapy or other precision medicine applications which are typically based on scarce biological specimens (e.g., serum samples [18]).

The Maver-1 and HEK293 triplicate DAA and DIA analyses and the respective transition lists have been deposited in the MassIVE repository (MassIVE dataset MSV000081439) under the creative commons zero license (CC0 1.0). The dataset is publicly available under the following link: <ftp://massive.ucsd.edu/MSV000081439>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Camilla Bacci (Philogen SpA) for providing W6/32 antibody.

Funding: This work was supported financially by ETH Zürich, the Swiss National Science Foundation, the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 305309 (PRIAT) and no. 305608 (EUREnOmics) and by the European Research Council (ERC advanced grant "ZAUBERKUGEL").

Abbreviations

HLA	human leukocyte antigen
DDA	data-dependent acquisition
DIA	data-independent acquisition
PSM	peptide-to-spectrum match

References

- [1]. Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol.* 2011; 11:823–836. [PubMed: 22076556]
- [2]. Schumacher FR, Delamarre L, Jhunjhunwala S, Modrusan Z, et al. Building proteomic tool boxes to monitor MHC class I and class II peptides. *Proteomics.* 2017; 17
- [3]. Caron E, Kowalewski DJ, Chiek Koh C, Sturm T, et al. Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry. *Mol Cell Proteomics.* 2015; 14:3105–3117. [PubMed: 26628741]
- [4]. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science.* 1992; 255:1261–1263. [PubMed: 1546328]
- [5]. Bassani-Sternberg M, Braunlein E, Klar R, Engleitner T, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun.* 2016; 7:13404. [PubMed: 27869121]
- [6]. Rotzschke O, Falk K, Deres K, Schild H, et al. Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature.* 1990; 348:252–254. [PubMed: 1700304]
- [7]. Bruderer R, Bernhardt OM, Gandhi T, Miladinovic SM, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics.* 2015; 14:1400–1410. [PubMed: 25724911]
- [8]. Caron E, Espona L, Kowalewski DJ, Schuster H, et al. An open-source computational and data resource to analyze digital maps of immunopeptidomes. *Elife.* 2015; 4
- [9]. Schittenhelm RB, Sivanesswaran S, Lim Kam Sian TC, Croft NP, Purcell AW. Human Leukocyte Antigen (HLA) B27 Allotype-Specific Binding and Candidate Arthritogenic Peptides Revealed through Heuristic Clustering of Data-independent Acquisition Mass Spectrometry (DIA-MS) Data. *Mol Cell Proteomics.* 2016; 15:1867–1876. [PubMed: 26929215]
- [10]. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods.* 2004; 1:39–45. [PubMed: 15782151]
- [11]. Hu A, Noble WS, Wolf-Yadlin A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res.* 2016; 5

- [12]. Zamo A, Ott G, Katzenberger T, Adam P, et al. Establishment of the MAVER-1 cell line, a model for leukemic and aggressive mantle cell lymphoma. *Haematologica*. 2006; 91:40–47. [PubMed: 16434369]
- [13]. Simmons NL. Tissue culture of established renal cell lines. *Methods Enzymol*. 1990; 191:426–436. [PubMed: 2074770]
- [14]. Dellgren C, Nehlin JO, Barington T. Cell surface expression level variation between two common Human Leukocyte Antigen alleles, HLA-A2 and HLA-B8, is dependent on the structure of the C terminal part of the alpha 2 and the alpha 3 domains. *PLoS One*. 2015; 10:e0135385. [PubMed: 26258424]
- [15]. Ritz D, Gloger A, Weide B, Garbe C, et al. High-sensitivity HLA class I peptidome analysis enables a precise definition of peptide motifs and the identification of peptides from cell lines and patients' sera. *Proteomics*. 2016; 16:1570–1580. [PubMed: 26992070]
- [16]. Rasmussen M, Harndahl M, Stryhn A, Boucherma R, et al. Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J Immunol*. 2014; 193:4790–4802. [PubMed: 25311805]
- [17]. Tsou CC, Avtonomov D, Larsen B, Tucholska M, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods*. 2015; 12:258–264. [PubMed: 25599550]
- [18]. Ritz D, Gloger A, Neri D, Fugmann T. Purification of soluble HLA class I complexes from human serum or plasma deliver high quality immuno peptidomes required for biomarker discovery. *Proteomics*. 2017; 17

Figure 1. DIA workflow for HLA peptidome analysis and method optimization.

(A) Schematic representation of the DIA workflow for HLA peptidome analysis. Cell lysates were subjected to HLA class I affinity purification and eluted peptides were analyzed by DDA and DIA. Spectral libraries were created with SpectraST from DDA data after filtering SEQUEST results for 1% FDR with Percolator. Transition lists were created from spectral libraries with spectrast2tsv and used to interrogate DIA data with Skyline. All detected peaks were reintegrated with mProphet and filtered to a q-value below 0.01. (B) Schematic representation of the different DIA methods. Each rectangle (light or dark grey)

represents one DIA window. Rectangles are drawn to scale. **(C)** Distribution of precursor masses from triplicate DDA analysis of the Maver-1 HLA peptidome. The m/z distributions of all peptides with charge state 2 ($z=2$, dark grey) and 3 ($z=3$, light grey) were plotted. **(D)** Comparison of DIA methods with different m/z ranges. Maver-1 HLA peptidomes were analyzed with DIA methods from 250-850 m/z or 400-700 m/z (see **B**). The difference in the number of peptides identified within a certain m/z range was plotted. **(E)** Comparison of DIA methods with windows of variable or fixed width. Maver-1 HLA peptidomes were analyzed with DIA methods with windows of either variable (250-850 m/z var) or fixed (250-850 m/z fix) width (see **B**). The difference in the number of peptides identified within a certain m/z range was plotted.

Figure 2. Characteristics of the Maver-1 HLA class I peptidome following DDA or DIA analysis. (A) Reproducibility of the Maver-1 HLA peptidome following DDA and DIA. HLA class I peptides were purified from 10^8 Maver-1 cells in triplicates each. Left: Three replicates were acquired in DDA mode and peptide identifications were filtered to 1% FDR after a database search with SEQUEST. Right: Three replicates were acquired in DIA mode and data was analyzed with a spectral library generated from the DDA samples using Skyline. The FDR after Skyline was estimated with mProphet and set to 1% at peptide precursor level. Euler-Venn diagrams display the overlap of peptide identifications between replicates. (B)

Reproducibility of peptide intensities between replicate DIA peptidome analyses. Intensities (i.e. Skyline total area) of peptides identified in two Maver-1 replicates were log₂-transformed and plotted. The linear correlation is plotted as dotted line, the corresponding R² is reported in the lower right corner. **(C)** Dynamic range of peptide identifications in DIA peptidome analysis. Intensities (i.e. Skyline total area) of peptides identified from a triplicate analysis of the Maver-1 peptidome are plotted. **(D)** NetMHCpan 3.0 binding prediction of 8 to 11-mers and assignment to corresponding HLA alleles. Peptides were annotated as predicted to bind if the rank received from NetMHCpan was below 2%. Predicted binders were assigned to the respective allele. **(E)** Evaluation of the scalability of the DIA peptidome analysis concerning the input cell number. HLA peptides were isolated from Maver-1 lysate corresponding to 1, 5, 10, 25 and 100 million Maver-1 cells. Samples were acquired in DIA mode and data was analyzed using Skyline. Peptides identified with an mProphet q-value below 0.01 are reported. **(F)** Effect of library size on peptide identification. Maver-1 HLA class I peptidomes were acquired in DIA mode in triplicate and analyzed using Skyline with spectral libraries generated from 3, 6, 12 or 16 Maver-1 HLA class I DDA analyses. The graph shows the proportion of peptides identified from the four libraries with an mProphet q-value below 0.01. Identified peptides are depicted in light grey, while library peptides not identified are depicted in dark grey. **(G)** Effect of library size on reproducibility. Maver-1 HLA class I peptidomes were acquired in triplicate and analyzed using Skyline with spectral libraries generated from 6 (left), 12 (middle) or 16 (right) DDA analyses. Euler-Venn diagrams display the overlap of peptide identifications between replicates.

Figure 3. Characteristics of the HEK293 HLA class I peptidome following DDA or DIA analysis. (A) Comparison of DIA methods with variable and fixed window widths. HEK293 HLA peptidomes were analyzed with DIA methods with 20 windows of either variable (250-850 m/z var) or fixed (250-850 m/z fix) width (see Figure 1B). The difference in the number of peptide identifications with a certain m/z range was plotted. (B) Distribution of precursor masses from a DDA analysis of the HEK293 HLA peptidome. DDA data was searched with SEQUEST and filtered to 1% FDR with Percolator. The m/z distributions of all peptides with charge state 2 ($z=2$, dark grey) and 3 ($z=3$, light grey) were plotted. (C) Reproducibility

of DDA and DIA data of the HEK293 HLA class I peptidome. HLA class I peptides were purified from 10^8 HEK293 cells. Three replicates were acquired in DDA mode, data was searched with SEQUEST and filtered to 1% FDR with Percolator. Three replicates were acquired in DIA mode and data was analyzed using Skyline. The FDR after Skyline was estimated with mProphet and set to 1% at peptide precursor level. Euler-Venn diagrams display the overlap of peptide identifications between replicates. **(D)** Reproducibility of peptide intensities between replicate DIA peptidome analyses. Intensities (i.e. Skyline total area) of peptides identified in two HEK293 replicates were $\log_2$ -transformed and plotted. The linear correlation is plotted as dotted line, the corresponding R^2 is reported in the lower right corner. **(E)** NetMHCpan 3.0 binding prediction of 8 to 11-mers and assignment to corresponding HLA alleles. Peptides were annotated as predicted to bind if the rank received from NetMHCpan was below 2%. Predicted binders were assigned to the respective allele. **(F)** Effect of library size on peptide identification. HEK293 HLA class I peptidomes were acquired in triplicate and analyzed with spectral libraries generated from 3, 6, 12 or 24 HEK293 HLA class I DDA analyses using Skyline. The graph shows the proportion of peptides that were identified from the five libraries with an mProphet q-value below 0.01. Identified peptides were depicted in light grey, while library peptides not identified are depicted in dark grey. **(G)** Effect of library size on reproducibility. HEK293 HLA class I peptidomes were acquired in triplicate and analyzed using Skyline with spectral libraries generated from 6 (left), 12 (middle) or 24 (right) DDA analyses. Euler-Venn diagrams display the overlap of peptide identifications between replicates.