Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD)

Geert P. M. Mommen[a,b,c], Christian K. Frese[b,c], Hugo D. Meiring[a], Jacqueline van Gaans-van den Brink[d], Ad P. J. M. de Jong[a], Cécile A. C. M. van Els[d], and Albert J. R. Heck[b,c,1]

[a]Formulation and Analytical Research, Institute for Translational Vaccinology, 3721 MA, Bilthoven, The Netherlands; [b]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH, Utrecht, The Netherlands; [c]Netherlands Proteomics Centre, 3584 CH, Utrecht, The Netherlands; and [d]Center for Immunology of Infectious Diseases and Vaccines, National Institute for Public Health and the Environment, 3721 MA, Bilthoven, The Netherlands

Edited by Peter Cresswell, Yale University School of Medicine, New Haven, CT, and approved February 13, 2014 (received for review November 15, 2013)

The identification of peptides presented by human leukocyte antigen (HLA) class I is tremendously important for the understanding of antigen presentation mechanisms under healthy or diseased conditions. Currently, mass spectrometry-based methods represent the best methodology for the identification of HLA class I-associated peptides. However, the HLA class I peptide repertoire remains largely unexplored because the variable nature of endogenous peptides represents difficulties in conventional peptide fragmentation technology. Here, we substantially enhanced (about threefold) the identification success rate of peptides presented by HLA class I using combined electron-transfer/higher-energy collision dissociation (EThcD), reporting over 12,000 high-confident (false discovery rate <1%) peptides from a single human B-cell line. The direct importance of such an unprecedented large dataset is highlighted by the discovery of unique features in antigen presentation. The observation that a substantial part of proteins is sampled across different HLA alleles, and the common occurrence of HLA class I nested sets, suggest that the constraints of HLA class I to comprehensively present the health states of cells are not as tight as previously thought. Our dataset contains a substantial set of peptides bearing a variety of posttranslational modifications presented with marked allele-specific differences. We propose that EThcD should become the method of choice in analyzing HLA class I-presented peptides.

human leukocyte antigen class I | electron-transfer dissociation | major histocompatibility complex | phosphorylation | binding motif

Class I molecules of the human leukocyte antigen (HLA) complex present short peptides, typically 8–11 aa in length at the cell surface, for scrutiny by the immune system (1). These peptide fragments are generated in the cytoplasm by proteasomal degradation of source proteins, translocated into the endoplasmic reticulum (ER) and subjected to N-terminal trimming to a size that is suitable for loading onto the HLA (2). Loading is governed by physicochemical binding motifs typical for each HLA class I allele (3). Depending on the motif required for the HLA class I allele(s) expressed, an ER-residing peptide may become presented or not. Recognition of specific HLA class I peptide complexes by CD8 T lymphocytes on pathogen infected or cancerous cells leads to the activation of a cytotoxic response and the clearance of the diseased cell. The identification of these HLA class I-associated peptides has important consequences for understanding the biology of cells, vaccine design, and tumor immunotherapy (4, 5).

Today mass spectrometry (MS) is the core technology for the analysis of HLA class I-presented peptides. These peptides are typically enriched from cell lysates through the affinity purification of HLA class I peptide complexes, released from the HLA by acid elution, and separated by liquid chromatography (LC) before introduction into the mass spectrometer. Identification is commonly accomplished by MS sequencing using collision-induced dissociation (CID) or beam-type higher-energy CID (HCD) (6). Both methods generate the peptide fragment ions that can be used for sequence identification in automated database search strategy or de novo sequence analysis. These methods have been thoroughly optimized and work particularly well for tryptic peptides that are produced by in vitro digestion of proteomes. However, for endogenously processed peptides, such as HLA class I-associated peptides, only a small fraction of the acquired tandem MS spectra (MS/MS) contains sufficient sequence-diagnostic information for correct assignment of the peptide sequence (6). To improve peptide identifications, the alternative fragmentation method electron-transfer dissociation (ETD) can be used, which complements CID particularly for longer and more basic peptides (7, 8). However, both CID and ETD may suffer from the limited sequence information concealed in the short HLA class I-associated peptide sequences and incomplete peptide fragmentation due to the occurrence of certain amino acid residues that are known to hamper efficient backbone dissociation (7). As a consequence, we hypothesized that a treasure of peptides presented by HLA class I molecules might still be in oblivion for identification by current MS technology.

Recently, we introduced a novel fragmentation scheme termed "electron-transfer/higher-energy collision dissociation" (EThcD) (9, 10). This method employs dual fragmentation after the

Significance

The surface presentation of peptides by human leukocyte antigen (HLA) class I molecules is essential for adaptive immune responses against pathogens and tumors. To date, the repertoire of HLA class I peptides remains largely unexplored, although their identity could lead to new targets for vaccine development or cancer immunotherapy. Here, we demonstrate that recent developments in mass spectrometry (MS)-based sequencing technology can expand the detectable peptide repertoire to an unprecedented depth, revealing unique features in the antigen presentation machinery. Our data include a variety of posttranslational modifications for which evidence is accumulating that they play important roles in human diseases. Hence, next-generation MS-based sequencing can make the important step into the discovery of disease-related HLA class I antigens.

Author contributions: G.P.M.M., A.P.J.M.d.J., and A.J.R.H. designed research; G.P.M.M. and C.K.F. performed research; C.K.F., H.D.M., J.v.G.-v.d.B., C.A.C.M.v.E., and A.J.R.H. contributed new reagents/analytic tools; G.P.M.M. and C.K.F. analyzed data; and G.P.M.M., C.K.F., H.D.M., A.P.J.M.d.J., C.A.C.M.v.E., and A.J.R.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: a.j.r.heck@uu.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321458111/-/DCSupplemental.

Downloaded at Palestinian Territory, occupied on December 29, 2021

IMMUNOLOGY

CHEMISTRY

www.manaraa.com

isolation of a single ion package to generate both the fragment ions induced by ETD (c/z) and HCD (b/y) in a single spectrum. The generation of dual-fragment ion series results in more informative MS/MS spectra that enable highly confident peptide assignment and localization of posttranslational modifications (PTMs) (9, 10). In this study, we explore the power of dual fragmentation EThcD for the analysis of peptides presented by HLA class I molecules on human B-lymphoblastoid cells. Our results revealed a supreme performance of EThcD in comparison with all established methods (CID, HCD, and ETD), resulting in an unprecedented inventory of HLA class I-associated peptides.

Results

Experimental Approach for the Identification of HLA Class I-Associated Peptides. To evaluate the full and unbiased potential of EThcD for the analysis of the HLA class I-associated peptides, we analyzed a complex repertoire of peptides presented by various class I molecules on the surface of an HLA-A, -B and -C heterozygous B-lymphoblastoid cell line, GR. HLA class I peptide complexes were isolated from a GR lysate by affinity purification, bound peptides were released by acid elution and analyzed by reversed-phase LC-MS/MS (11). This 1D strategy was used to assess the performance of all possible peptide fragmentation techniques currently available on the Orbitrap Elite instrument (Thermo Fisher Scientific) used here. To evaluate whether sequencing-short, endogenously processed peptides benefits from the combined information generated by two different fragmentation modes, we compared EThcD fragmentation with sequential CID/HCD (in silico spectral merging) or single fragmentation methods CID, HCD, and ETD (*SI Appendix*, Fig. S1). For each fragmentation technique, fragment ion spectra (MS/MS) were acquired with high-mass accuracy and high resolution using the Orbitrap analyzer, not only to ensure consistency between the datasets, but also to improve specificity in the database search analysis. MS/MS spectra were searched against the human proteome using SEQUEST (Thermo Fisher Scientific) and filtered to a <1% false discovery rate (FDR) using the percolator algorithm (12). We found that percolator performed approximately twofold better than a standard target-decoy approach (13), irrespective of the fragmentation method used.

EThcD Boosts the Identification of HLA Class I-Associated Peptides. The peptide identification results from the 1D strategy are summarized in Table 1. Although the number of MS/MS spectra were roughly identical (~12,000), the identification success rate and the number of uniquely identified peptides clearly indicate that EThcD fragmentation by far outperforms all other fragmentation techniques. We found that 39% of the MS/MS attempts by EThcD lead to high-confident peptide assignments, which is approximately a factor of 3 higher in comparison with HCD and ETD alone. The excellent performance of EThcD can be directly attributed to the more extensive backbone fragmentation and the generation of information-rich MS/MS spectra. As exemplified in Fig. 1, dissociation of the HLA-B27 restricted peptide YRAPELLL by HCD and ETD, respectively, generated limited sequence information, primarily due to the near-complete absence of fragment ions derived from the peptide C terminus. Dual fragmentation EThcD generated the complementary c/z and b/y ions, enabling high-confidence sequence assignment. The improvement in spectral quality is also reflected in the SEQUEST Xcorr distributions of the complete dataset, as significant higher scores were found for EThcD due to the assignment of both c/z and b/y ions (*SI Appendix*, Fig. S2). These findings were sustained by sequential CID/HCD, which performed significantly better than single stage fragmentation (CID and HCD alone). The identification of 23 arbitrarily selected peptides was validated using synthetic peptide counterparts. This analysis revealed an excellent spectral match between the endogenous peptides and the synthetic peptide counterparts, and also confirmed the enhancement in spectral richness by EThcD (*SI Appendix*, Table S4 and Figs. S10 and S11).

To benchmark and compare these results, we analyzed a tryptic *Escherichia coli* digest (5 ng) and found that the identification success rate of EThcD for sequencing endogenous peptides agrees well with the success rate for low-abundant tryptic peptides (39% versus 30%) (*SI Appendix*, Table S1). However, in contrast to endogenous peptides, for tryptic peptides the success rate of EThcD and HCD is comparable (30% and 34%, respectively), which is in agreement with previous data (9, 10). Together, the data unambiguously demonstrate that sequencing short, endogenous peptides is rather cumbersome when using standard HCD or ETD, and benefits considerably when combining the complementary sequence information induced by EThcD.

Global Mapping of the HLA Class I Peptide Repertoire. To more comprehensively profile the repertoire of peptides presented by HLA class I, we next used a 2D peptide separation strategy using strong cation exchange (SCX) fractionation followed by LC-MS/MS analysis. Here, we focused on exclusively using EThcD and sequential CID/HCD because these two methods performed best based on the initial 1D strategy. In total, 9,015 and 6,381 unique 8- to 14-mer peptides were identified by EThcD and sequential CID/HCD, respectively (Table 1). Note that generating high-quality MS/MS spectra by EThcD comes at the cost of a slightly slower duty cycle, thus leading to undersampling if complex samples are not fractionated. The most optimal balance between peptide identification coverage and the required analysis time was therefore found by the 2D strategy using EThcD, covering 74% of the complete set of unique HLA class I peptides (Fig. 2*D*).

To validate whether the identified peptides were indeed initially bound to HLA molecules, we predicted the HLA class I binding affinities using the NetMHC algorithm (14). Importantly, 90% of the identified peptides were predicted as strong binders (<1,000 nM IC_{50}) to a single HLA-A and -B allele on GR cells, which further strengthens our confidence in the

Table 1. Summary of data obtained in the LC-MS/MS analysis of HLA-associated peptides presented by GR B cells

	1D strategy					2D strategy		
Dataset	EThcD	CID/HCD	CID	HCD	ETD	EThcD	CID/HCD	Combined data
MS/MS spectra	11,777	10,214*	13,394	16,054	11,223	39,156	42,227*	196,485
PSM, <1% FDR	4,532	2,418	1,319	1,948	2,545	13,555	8,794	35,111
Identification rate, %	39	12†	10	12	23	35	10†	19
Unique peptides	3,454	2,142	1,273	1,737	2,288	9,255	6,554	12,699
Unique peptides 8–14 aa	3,381	2,027	1,174	1,622	2,215	9,015	6,381	12,199
No. of proteins	2,205	1,517	982	1,268	1,650	4,536	3,706	5,603
Analysis time, h	3	3	3	3	3	9 × 2	9 × 2	51

PSM, peptide-spectrum match.
*The number of merged CID/HCD spectra used for database search analysis.
†Identification rate was calculated from twice the number of merged CID/HCD spectra.

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

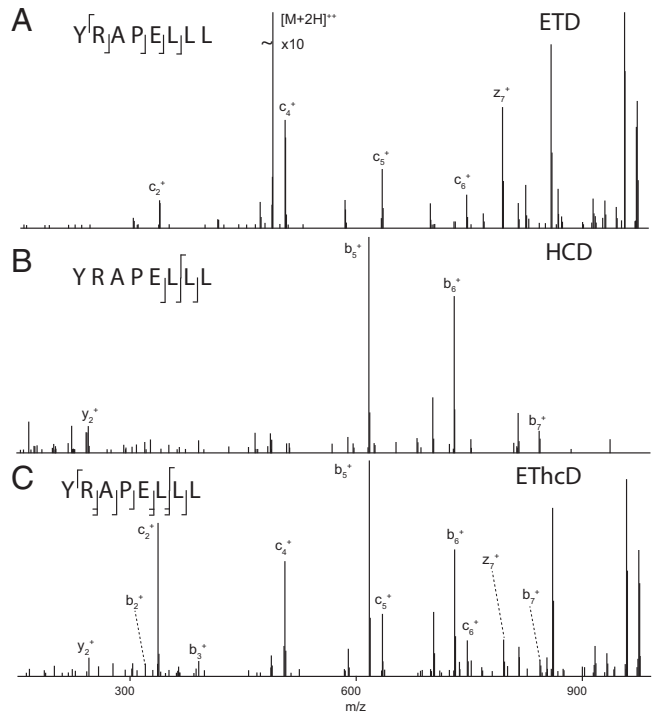

Fig. 1. Peptide fragmentation by ETD, HCD, and EThcD. Illustrative MS/MS spectra of the HLA-B27-associated peptide YRAPELLL upon fragmentation by (*A*) ETD, (*B*) HCD, and (*C*) EThcD. The observed and/or assignable c/z and b/y fragment ions are indicated above and below the peptide sequence. MS/MS of the synthetic peptide counterpart are provided in *SI Appendix*, Fig. S11.

correctness of the peptide assignments. Using sequence-based rules on the cumulative dataset, the NetMHC algorithm assigned 1,170 peptides to HLA-A1 (9%), 2,132 peptides to HLA-A3 (17%), 4,314 peptides to HLA-B7 (35%), and 3,530 peptides to HLA-B27 (29%), whereas 1,234 peptides could not be classified on these terms [not assigned (NA)]. The NA peptides include besides weak predicted binders also a low contribution of peptides bound to the HLA-C alleles expressed (*SI Appendix*, Fig. S3) (15), but we found their numbers relative to the total assigned repertoire too low to further specify in this study.

Comparison of the complete dataset with recent large-scale studies revealed that 81% of the peptides have not been reported before (*SI Appendix*, Table S2) (16–18). Remarkably, the unique peptides identified in this study have distinct sequence characteristics compared with literature data (*SI Appendix*, Fig S4). Sequence logos of the uniquely identified peptides here revealed overrepresentation of basic residues (Arg, Lys) at the peptide N terminus (P1 or P3 position) for HLA-A3, -B7, and -B27, indicating that these peptides have been underrepresented in previous studies.

Fragmentation-Related Bias in Profiling the HLA Class I Peptide Repertoire. To test for fragmentation-related biases, we further specified the identification results for each of the applied fragmentation method with respect to the different HLA class I alleles (Fig. 2). This evaluation included, for each HLA allele, the number of uniquely assigned peptides (Fig. 2*A*), the peptide binding consensus (Fig. 2*B*), and the overlap in peptide identification between the fragmentation techniques (Fig. 2*C*). The observed sequence logos clearly matched the known binding motifs of the HLA alleles, through amino acids anchor residues at position P2 (P3) and the C-terminal amino acid (PΩ). EThcD provided the largest number of peptide identifications across all HLA alleles (Fig. 2*A*). The unmatched performance of EThcD was most pronounced for the HLA-B7 and -B27. The peptides bound to these alleles are more difficult to sequence by conventional methods due to the preferred presence of a proline or an arginine residue at the P2 position (7). A known limitation of ETD is the inability to cleave the N-Cα bond of proline, whereas CID/HCD suffers from inefficient peptide dissociation due to internal arginine residues that are known to hamper random backbone protonation. The generation of dual-ion series in EThcD overcomes these limitations, resulting in an almost twofold improvement in the identification of peptides bound to HLA-B7 and -B27.

To compare the performance of EThcD with the conventional fragmentation techniques in more detail, we used the 1D dataset to determine the normalized overlap between the peptide identification results for each HLA allele. Fig. 2*C* shows that the added value of the conventional fragmentation techniques is limited because EThcD accounts for ∼80% of the peptide identifications across all HLA alleles. It is worth mentioning, but not surprising, that ETD alone performed particularly well for these HLA-B27–bound peptides. This was reflected by the absolute number of identified peptides (Fig. 2*A*), as well as the complementary advantage of ETD for sequencing of triply charged peptides (*SI Appendix*, Fig. S5). These findings are in line with previous reports describing that peptides with a strong basic N terminus generate more straightforward and easy-to-interpret MS/MS spectra in ETD (19).

Characteristics of the HLA Class I Peptide Repertoire. The general characteristics of the HLA class I peptide repertoire identified here add to existing knowledge on self-peptidome. The 12,199 unique peptides originate from 5,603 source proteins, which are primarily located in the ER, nucleus, and cytoplasm. Distribution analysis of the number of peptides identified per protein revealed that 48% (2,636 proteins) were represented by a single peptide, whereas the remaining proteins were represented by multiple unique peptides (*SI Appendix*, Fig. S6). As reported previously (16), the HLA class I sampling rate seems to correlate with protein length (*SI Appendix*, Fig. S7). For proteins that were represented by multiple peptides, we found that 2,295 proteins were represented by at least two different HLA alleles (*SI Appendix*, Fig. S6). Moreover, for the peptides related to these proteins, no relation was found between peptide intensity, allele specificity, or predicted binding strength, confirming that the final level of presentation depends on many factors [e.g., proteolytic activity, transporter-associated with antigen processing (TAP) transport efficiency] (16, 20).

HLA class I molecules predominantly present peptides of 9–11 aa in length. However, there are numerous reports of longer peptides that can also elicit cytotoxic T-cell responses, and some of these unusually long peptides fully overlap in sequence with their counterparts that have more conventional length (21). In our dataset we observed 1,376 peptides pairs with an identical core sequence but having 1- to 5-aa-length differences. These additional amino acids were detected either at the N terminus ($n = 341$), the C terminus ($n = 829$), or with no consensus at their termini (i.e., ragged ends at both termini; $n = 206$). Mapping these nested peptide species to different HLA alleles revealed unexpected differences between the frequency of N- and C-terminal–extended peptide pairs (*SI Appendix*, Fig. S8). For HLA-B alleles, significantly more C-terminal–extended peptide pairs were detected compared with N-terminal–extended pairs, as reported similarly for HLA-B27–bound peptides (17). In contrast to these findings in HLA-B, a more equal frequency distribution between N- and C-terminal–extended length variants was found for HLA-A. Frequency analysis of the C-terminal anchor residues revealed stricter constraints on the C-terminal anchor residue of the HLA-A versus the HLA-B alleles (*SI Appendix*, Fig. S8). For example, the C terminus of HLA-A1–bound peptide pairs are predominated by a tyrosine residue, although there is less restriction for variants bound to HLA-B27

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

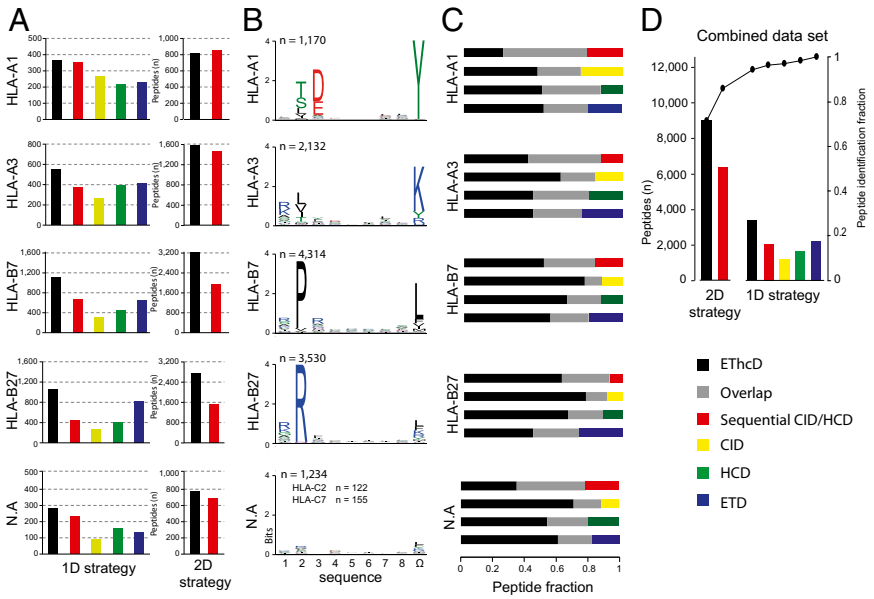

Fig. 2. Performance characteristics of peptide fragmentation techniques for HLA class I-associated peptides. Shown are identification results and binding motifs of peptides presented by different HLA class I molecules on the surface of GR cells. (*A*) The number of unique 8- to 14-mer peptides identified by the 1D and 2D peptide separation strategy, assigned to HLA-A1, -A3, -B7, and -B27 molecules, or not assigned to any of the above (NA). The 1D strategy comprised consecutive single LC-MS/MS analysis runs, using either EThcD (black), sequential CID/HCD (red), CID (yellow), HCD (green), or ETD (blue). The 2D strategy included the analysis of SCX fractions by LC-MS/MS using EThcD (black) or sequential CID/HCD (red). (*B*) Sequence logos showing that peptides are bound to a particular HLA molecule through a distinct binding motif, typically by anchor residues on position P2 (P3) and the C-terminal PΩ residue. (*C*) Bar diagrams depicting the peptide fraction identified by EThcD (black), the overlap between the methods (gray), and uniquely identified by CID/HCD (red), CID (yellow), HCD (green), or ETD (blue). (*D*) The number of unique peptides identified by the different peptide fragmentation methods (color coded) in the 1D and 2D strategy. The line graph depicts the cumulative number of unique HLA class I-associated peptides (left *y* axis) and the peptide identification fraction (right *y* axis).

(Fig. 2*B*). For HLA class I nested sets, no relation was found between peptide intensity, predicted binding strength, or peptide length.

PTM HLA Class I Peptides. Peptides presented by HLA molecules can harbor PTMs (22). Table 2 summarizes the PTMs identified on HLA class I-associated peptides and their unmodified counterparts. The total numbers of unmodified counterparts are lower compared with PTM peptides. This was primarily attributed to undersampling, although for some PTMs the modification reaction may proceed quantitatively (i.e., cysteinylation), limiting the detection of unmodified counterparts.

The frequently observed asparagine deamidation reaction here can proceed either spontaneously (aging) or enzymatically. It has been demonstrated that N-glycosylated peptides may be presented by HLA class I molecules, but only after back transport of ER proteins to the cytosol, with concomitant enzymatic deglycosylation, resulting in the deamidation of the glycosylated asparagine (23). Approximately 45% of the deamidated peptides were derived from membrane-associated glycoproteins with the correct site-specific glycosylation motifs (*SI Appendix*, Fig. S9). Interestingly, for HLA-A1 only, we found a relatively high proportion of deamidated peptides in which the asparagine residue on position 3 was converted to an aspartic acid, the anchor residue for this class I allele (*SI Appendix*, Table S3). For example, the LSNISHLNY sequence of glycosyltransferase BGnT-2 was only detected in the deamidated form LSDISHLNY.

We also detected a total of 59 serine or threonine phosphorylated peptides. Phosphorylated peptides were assigned to both HLA-A and -B alleles at similar (normalized) frequencies (*SI Appendix*, Table S3). A preferred bias was found for a basic residue on P1 and the phosphate group on P4 of peptide sequence (*SI Appendix*, Fig. S9)—characteristics similar to those previously reported (24, 25). Mapping the data to phosphorylation-specific motifs that are commonly associated with specific kinases resulted in the assignment of 72% of the phosphopeptides to a proline-directed kinase motif (PxTP, PxSP), similar to those reported by Cobbold et al. (24), or basophilic kinase motif (RxxS) (*SI Appendix*, Fig. S9). Comparison with previous studies revealed that 16 HLA-B7–associated phosphopeptides had been reported previously (16, 24, 26), whereas a total of 24 phosphorylation sites are annotated in the Uniprot protein database (www.uniprot.org/). The correct identification of selected phosphopeptides could be nicely confirmed by the spectra we generated from their synthetic peptide counterparts (*SI Appendix*, Fig. S10).

Cysteinylation is a spontaneous reaction where a cysteine forms a disulfide bond to a free cysteine molecule, either in vivo or in vitro (27, 28). In assigning the 196 cysteinylated peptides to the HLA-A and -B alleles, we unexpectedly found a bias toward the HLA-A1. Compared with unmodified peptides, 6% of the HLA-A1–associated peptides were cysteinylated, whereas the frequency of this modification was only 1–2% for the remaining alleles (*SI Appendix*, Table S3).

Discussion

Identification of HLA class I-associated peptides by MS has contributed substantially to the knowledge of the antigen-processing and presentation mechanisms involved in cellular immune response (6). However, the identity of a considerable part of the repertoire of peptides presented by HLA class I molecules remains largely unknown due to current limitations in MS sequencing. The immediate interest of this study was therefore to uncover peptides that escaped identification by currently used sequencing methods CID, HCD, and ETD. Therefore, we applied EThcD fragmentation for the identification of HLA class I-associated peptides, a dual fragmentation method that generates ETD- and HCD-derived fragment ions from a single ion package for enhanced peptide fragmentation and sequence identification (9).

The most important finding of this comprehensive evaluation study was the excellent performance of EThcD for the global analysis of the HLA peptide repertoire, revealing a larger HLA class I peptidome for a single cell line than previously reported

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

Table 2. Overview of HLA-associated peptides harboring PTMs

PTMs	Modified peptides	Unmodified counterpart	Uniprot annotated	Preferred position (%)
Asn deamination	83	47	26*	P3–P6 (76)
Ser phosphorylation	44	24	22	P4–P7 (78)
Thr phosphorylation	15	6	2	P4 (67)
Met oxidation	988	725		
Cysteinylation	196	12		
N-cyclization (pyroQ)	246	166		

pyroQ, pyrrolidone carboxylic acid.
*Annotated N-glycosylation sites.

(15). Strikingly, almost 40% of the MS/MS spectra generated by EThcD led to highly confident assignments of peptide sequences, which is, to our knowledge, the most advanced peptide identification success rate for endogenous peptides to date. EThcD generates more extensive backbone fragmentation and increased coverage of the peptide sequence, resulting in more readily identified HLA class I-associated peptides. In comparison, Hassan et al. (16) recently reported a large inventory of HLA class I peptides using CID-based MS, but implemented less strict confidence criteria (FDR of <10% instead of <1% here). The advantage of EThcD in generating more informative MS/MS spectra for high-confident peptide assignments was supported by the in silico spectral merging of separately acquired CID and HCD spectra. Contrary to endogenous peptides, EThcD performs equally as well as HCD for tryptic peptides generated by in vitro digestion of proteomes (9). This discrepancy is primarily attributed to the fact that endogenous peptides are more difficult to sequence by HCD, whereas current MS technology and database search algorithms work particularly well for tryptic peptides and do not require improvements in spectral quality just for their identification per se.

The selection of a B-lymphoblastoid cell line that was heterozygous for HLA-A, -B, and -C allowed us to evaluate the allele-specific performance of the different types of fragmentation techniques. A remarkable finding was that EThcD provided unbiased insight into the repertoire of presented peptides. In contrast, biases in allele-specific peptide identification results could be directly translated to the known limitations of conventional peptide fragmentation techniques. This suggests that the outcome of HLA class I peptide identification studies that are based on CID (HCD) (16, 18) or ETD (29) alone may only partially reflect the actual landscape of peptides presented at the cell surface. EThcD as the universal fragmentation method is therefore an attractive alternative to current MS sequencing technology.

The high-discovery rate in this study of peptides that constitute the HLA class I repertoire not only allowed us to confirm typical known features, but also to add unique findings. On a global scale, we assigned HLA class I peptides to ∼5,600 source proteins, which is in agreement with the findings of Hassan et al. (16). Considering that cultured human cells express ∼11,000 protein-coding genes (30), mapping the HLA class I peptidome is now becoming possible for 50% of the expressed proteins. Moreover, ∼40% of the source proteins are represented by at least two peptides bound to different HLA molecules. This broad representation of the cell's proteome by HLA class I is in agreement with the general belief that equitable sampling is necessary to comprehensively reflect the health status of the cell (31). However, although the major part of the proteome seemed suited for HLA sampling, the actual small number of peptides identified per source protein confirms that considerable selection of peptides takes place, related to proteolytic cleavage, efficiency of TAP and the binding affinity of the available HLA molecule (16, 20). Earlier studies reporting fewer numbers of peptides (hundreds) demonstrated that most proteins were represented by only a single peptide (18). Hence, fragmentation techniques that can capture the full landscape of HLA class I peptides in cells of interest, such as EThcD, can make an important step toward the next stage of discovery.

Our study furthermore revealed various allele- or peptide-specific features that can lead to a better understanding of the peptide processing and specificity in various HLA alleles. We identified a large set of nested peptides that differ in length either at the N or C terminus. The presence of N-terminal–extended length variants have been reported previously (17, 32). The activity of aminopeptidases in the ER effectively edits the majority of endogenously processed peptides to a length suitable for loading onto the HLA class I molecules, but a minority of peptides escape this efficient trimming process (33). Only a few studies report C-terminally extended peptide pairs. Ben Dror et al. (17) attributed these C-terminal–extended peptides to the broad cleavage specificity of the proteasome, whereas Lorente et al. (34) suggested the additional role of C-terminal trimming by carboxypeptidase. The global picture emerging from this study revealed an important role of allele-specific physicochemical binding motifs. The observed differential C-terminal length variation for the various HLA alleles here were directly attributed to the different constraints for C-terminal binding between the HLA alleles, in particular because we could assume equal enzymatic activities because the peptides were extracted from a single cell line.

In addition, detailed findings included marked differences in the number of PTM peptides presented by different HLA class I molecules. The presentation of certain HLA-A1–associated peptides is enhanced by the deamidation of asparagine into aspartic acid. This modification creates a negatively charged side chain with the ability to form hydrogen bonds and hence servers as a high-affinity anchor residue for HLA-A1. Such a mechanism has been previously reported for glutamine-containing peptides in HLA class II (35). The cysteinylation of cysteine residues is a nonenzymatic reaction which has been shown to play a role in T-cell recognition (27, 28). However, the biological relevance of this modification is rather unclear because this reaction can proceed in vivo or in vitro, and can both positively and negatively affect HLA binding (36). We found an altered (fourfold higher) frequency of cysteinylated peptides bound to HLA-A1 compared with the other alleles investigated in this study. This suggests that cysteinylation of peptides might occur before binding to the HLA molecule in the ER, and thus modulate the ability of peptides to bind to certain HLA alleles.

Finally, the presentation of HLA class I phosphopeptides and their recognition by T cells has increasingly been reported to play an important role in human diseases as the deregulation of signaling pathways is a hallmark of malignant transformation (24, 25). Although high-confident assignment of HLA class I phosphopeptides is challenging (16), we used the advantage of EThcD for unambiguous phosphosite localization (10) and validated our results using synthetic peptide counterparts and bioinformatics (kinase motifs). The fraction of phosphorylated peptides identified in this study is ∼0.5% of all peptides, irrespective of the alleles investigated. This suggests that there is no significant bias toward a certain HLA class I molecule for presenting phosphorylated peptides. The preference of a basic residue

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com

on P1 and the phosphate group on P4 of the sequence confirms previously reported roles of these biases in the preferential binding of phosphopeptides to HLA class I molecules (24, 25).

In summary, we present a large-scale analysis of HLA class I-associated peptides by EThcD fragmentation. The advantage of EThcD as a universal fragmentation technique was demonstrated by the improved number of identified peptides and the increased coverage of fragment ions over conventional fragmentation techniques. The data generated here further expand our knowledge of peptide presentation by HLA class I molecules. We therefore foresee a predominant role for EThcD in the comprehensive analysis of endogenously processed peptides, also beyond the HLA class I peptide repertoire.

Materials and Methods

The Cultured Cell Line and Isolation of HLA Class I-Associated Peptides. The HLA-A*01,-03, B*07,-27, and -C*02,-07–positive B-lymphoblastoid cell-line GR was grown in RPMI-1640 medium to a total number of 9×10^9 cells. HLA class I peptide complexes were immunoprecipitated from lysed GR cells (37), using the HLA-A–, -B–, and -C–specific mouse monoclonal IgG2a antibody W6/32. HLA class I-associated peptides were eluted with 10% (vol/vol) acetic acid and collected by passage over a 10-kDa molecular weight cutoff membrane.

Synthetic Peptides. Synthetic peptides were purchased from Pepscan.

MS. For 1D analysis, the HLA elution sample was analyzed directly by nanoscale LC-MS/MS using a Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific) and ETD-enabled LTQ Orbitrap Elite mass spectrometer (9). The system comprises a 20 × 0.1 mm i.d. trapping column (Reprosil C18, 3 μm; Dr. Maisch) and a 40 × 0.0075 cm i.d. analytical column (Zorbax SB-C18; 1.8 μm). Full MS spectra (*m/z* 300–1,500) were acquired in an Orbitrap at a resolution of 60,000 (FWHM). The 10 most abundant precursor ions were selected for either data-dependent EThcD, CID, HCD, or ETD fragmentation (unknown and 1+ charge state excluded). For sequential CID/HCD analysis, each precursor ion was sequentially selected for CID and HCD (top-five method). The maximum ion accumulation time for MS scans was set to 200 ms and for MS/MS scans to 2,500 ms. Fragment ions were detected in an Orbitrap at a resolution of 15,000 (FWHM). Dynamic exclusion was enabled with a repeat count of 1- and 60-s exclusion duration. The background ions at *m/z* 391.2843 and 445.1200 were used as lock mass. For the 2D strategy, HLA class I-eluted peptides were fractionated by SCX chromatography. The system comprises a Hypercarb trapping column (5 × 0.2 mm i.d., 7-μm particle size; Thermo Fisher) and SCX column (12 × 0.02 cm i.d. polysulfoethyl aspartamide, 5 μm; Poly LC). A total number of nine SCX fractions were subjected to LC-MS/MS, as described for 1D analysis.

Data Analysis. The analysis of MS raw data was carried out using Proteome Discoverer 1.4 software package (Thermo Fisher Scientific). The nonfragment filter was used to simplify ETD spectra and the spectrum grouper function to merge consecutive CID and HCD spectra. MS/MS scans were searched against the human Uniprot database (www.uniprot.org/) with no enzyme specificity using the SEQUEST HT mode. Precursor ion and MS/MS tolerances were set to 3 ppm and 0.02 Da, respectively. Asparagine deamidation, methionine oxidation, cysteinylation, phosphorylation (S, T, Y), or N-terminal glutamate cyclization were set as variable modifications. Results were filtered to <1% FDR using percolator (12), Xcorr >1.75, and rank = 1.

ACKNOWLEDGMENTS. G.P.M.M., C.K.F., A.P.J.M.d.J., and A.J.R.H. acknowledge support from the Netherlands Proteomics Centre. A.J.R.H. is additionally supported by the Institute for Chemical Immunology, a Nederlandse Organisatie voor Wetenschappelijk Onderzoek Gravitation project funded by the Ministry of Education, Culture and Science. C.K.F. and A.J.R.H. are partly financed by the European Community's Seventh Framework Programme (FP7/2007–2013) by PRIME-XS Project Grant Agreement 262067.

1. Parkin J, Cohen B (2001) An overview of the immune system. *Lancet* 357(9270): 1777–1789.
2. Neefjes J, Jongsma ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11(12):823–836.
3. Rammensee HG, Friede T, Stevanoviic S (1995) MHC ligands and peptide motifs: First listing. *Immunogenetics* 41(4):178–228.
4. Kessler JH, Melief CJ (2007) Identification of T-cell epitopes for cancer immunotherapy. *Leukemia* 21(9):1859–1874.
5. Ovsyannikova IG, Johnson KL, Bergen HR, 3rd, Poland GA (2007) Mass spectrometry and peptide-based vaccine development. *Clin Pharmacol Ther* 82(6):644–652.
6. Mester G, Hoffmann V, Stevanović S (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. *Cell Mol Life Sci* 68(9): 1521–1532.
7. Chi A, et al. (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci USA* 104(7):2193–2198.
8. Frese CK, et al. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res* 10(5):2377–2388.
9. Frese CK, et al. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* 84(22):9668–9673.
10. Frese CK, et al. (2013) Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res* 12(3):1520–1525.
11. Hunt DF, et al. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 255(5049):1261–1263.
12. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4(11):923–925.
13. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4(3):207–214.
14. Lundegaard C, et al. (2008) NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* 36(Web Server issue):W509–12.
15. Zhang H, Lundegaard C, Nielsen M (2009) Pan-specific MHC class I predictors: A benchmark of HLA class I pan-specific prediction methods. *Bioinformatics* 25(1):83–89.
16. Hassan C, et al. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. *Mol Cell Proteomics* 12(7):1829–1843.
17. Ben Dror L, Barnea E, Beer I, Mann M, Admon A (2010) The HLA-B*2705 peptidome. *Arthritis Rheum* 62(2):420–429.
18. Granados DP, et al. (2012) MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. *Blood* 119(26):e181–e191.
19. Altelaar AF, et al. (2012) Database independent proteomics analysis of the ostrich and human proteome. *Proc Natl Acad Sci USA* 109(2):407–412.
20. Yewdell JW (2006) Confronting complexity: Real-world immunodominance in antiviral CD8+ T cell responses. *Immunity* 25(4):533–543.
21. Burrows SR, Rossjohn J, McCluskey J (2006) Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol* 27(1):11–16.
22. Petersen J, Purcell AW, Rossjohn J (2009) Post-translationally modified T cell epitopes: Immune recognition and immunotherapy. *J Mol Med (Berl)* 87(11):1045–1051.
23. Skipper JC, et al. (1996) An HLA-A2-restricted tyrosinase antigen on melanoma cells results from posttranslational modification and suggests a novel pathway for processing of membrane proteins. *J Exp Med* 183(2):527–534.
24. Cobbold M, et al. (2013) MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. *Sci Transl Med* 5(203):203ra125.
25. Zarling AL, et al. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci USA* 103(40):14889–14894.
26. Meyer VS, et al. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. *J Proteome Res* 8(7): 3666–3674.
27. Meadows L, et al. (1997) The HLA-A*0201-restricted H-Y antigen contains a post-translationally modified cysteine that significantly affects T cell recognition. *Immunity* 6(3):273–281.
28. Pierce RA, et al. (1999) Cutting edge: The HLA-A*0101-restricted HY minor histocompatibility antigen originates from DFFRY and contains a cysteinylated cysteine residue as identified by a novel mass spectrometric technique. *J Immunol* 163(12): 6360–6364.
29. Ostrov DA, et al. (2012) Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. *Proc Natl Acad Sci USA* 109(25):9959–9964.
30. Nagaraj N, et al. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 7:548.
31. Hickman HD, et al. (2004) Toward a definition of self: Proteomic evaluation of the class I peptide repertoire. *J Immunol* 172(5):2944–2952.
32. Escobar H, et al. (2008) Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. *J Immunol* 181(7): 4874–4882.
33. Hearn A, York IA, Rock KL (2009) The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J Immunol* 183(9):5526–5536.
34. Lorente E, et al. (2013) Diversity of natural self-derived ligands presented by different HLA class I molecules in transporter antigen processing-deficient cells. *PLoS ONE* 8(3): e59118.
35. Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM (2004) Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci USA* 101(12):4175–4179.
36. Kittlesen DJ, et al. (1998) Human melanoma patients recognize an HLA-A1-restricted CTL epitope from tyrosinase containing two cysteine residues: Implications for tumor vaccine development. *J Immunol* 160(5):2099–2106.
37. van Els CA, et al. (2000) A single naturally processed measles virus peptide fully dominates the HLA-A*0201-associated peptide display and is mutated at its anchor position in persistent viral strains. *Eur J Immunol* 30(4):1172–1181.

Downloaded at Palestinian Territory, occupied on December 29, 2021
www.manaraa.com