© 2017 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

A Molecular Basis for the Presentation of Phosphorylated Peptides by HLA-B Antigens*[S]

Adán Alpízar‡§§, Fabio Marino§¶§§, Antonio Ramos-Fernández‖, Manuel Lombardía‡, Anita Jeko§¶, Florencio Pazos, Alberto Paradela‡, César Santiago‡‡, Albert J. R. Heck§¶ ¶¶, and Miguel Marcilla‡¶¶**

As aberrant protein phosphorylation is a hallmark of tumor cells, the display of tumor-specific phosphopeptides by Human Leukocyte Antigen (HLA) class I molecules can be exploited in the treatment of cancer by T-cell-based immunotherapy. Yet, the characterization and prediction of HLA-I phospholigands is challenging as the molecular determinants of the presentation of such post-translationally modified peptides are not fully understood. Here, we employed a peptidomic workflow to identify 256 unique phosphorylated ligands associated with HLA-B*40, -B*27, -B*39, or -B*07. Remarkably, these phosphopeptides showed similar molecular features. Besides the specific anchor motifs imposed by the binding groove of each allotype, the predominance of phosphorylation at peptide position 4 (P4) became strikingly evident, as was the enrichment of basic residues at P1. To determine the structural basis of this observation, we carried out a series of peptide binding assays and solved the crystal structures of HLA-B*40 in complex with a phosphorylated ligand or its nonphosphorylated counterpart. Overall, our data provide a clear explanation to the common motif found in the phosphopeptidomes associated to different HLA-B molecules. The high prevalence of phosphorylation at P4 is dictated by the presence of the conserved residue Arg62 in the heavy chain, a structural feature shared by most HLA-B alleles. In contrast, the preference for basic residues at P1 is allotype-dependent and might be linked to the structure of the A pocket. This molecular understanding of the presentation of phosphopeptides by HLA-B molecules provides a base for the improved prediction and identification of phosphorylated neo-antigens, as potentially used for cancer immunotherapy. ***Molecular & Cellular Proteomics 16: 10.1074/mcp.M116.063800, 181–193, 2017.***

Human leucocyte antigen (HLA)[1] class I molecules display at the cell surface peptide ligands derived from the degradation of endogenous proteins and present them to cytotoxic T lymphocytes (CTLs). In this way, virally infected or tumor cells can be specifically recognized by the immune system leading to the activation of CTLs that exert their cytotoxic effect on the antigen presenting cell. Classical HLA-I molecules -encoded by genes in the HLA-A, -B and -C loci- encompass over 7000 different proteins derived from more than 10,000 alleles (1). The peptide repertoire associated to a particular class I allotype includes several thousand ligands with defined structural motifs that allow their binding to the class I molecule. These peptides can also harbor post-translational modifications (PTMs) (2). In particular, phosphorylated HLA-I ligands have lately received attention, being proposed as potential targets for cancer immunotherapy on the basis that aberrant phosphorylation is a hallmark of tumor cells (3, 4). In this context, it is known that CTLs can recognize and respond specifically to phosphorylated HLA-I epitopes (5, 6).

Despite their potential relevance for cancer immunotherapy, the number of HLA class I-bound phosphopeptides described so far is limited and phosphopeptidomic studies have been performed for a few allotypes only. For instance, Zarling *et al.* identified in several lymphoblastoid cell lines and a lung carcinoma 15 phospholigands bound to different class I molecules, among them 6 presented by HLA-B*27 and 6 by HLA-B*07 (5). The same group also described a total of 36 HLA-A*02 phospholigands cumulatively presented by melanoma, ovarian carcinoma, and lymphoblastoid cells (3). Meyer

From the ‡Proteomics Unit, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain; §Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; ¶Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, the Netherlands; ‖Proteobotics SL, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain; **Computational Systems Biology Group, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain; ‡‡Macromolecular X-ray Crystallography Unit, Spanish National Biotechnology Centre (CSIC), Darwin 3, 28049, Madrid, Spain

Received October 3, 2016, and in revised form, November 16, 2016
Published, MCP Papers in Press, November 16, 2016, DOI 10.1074/mcp.M116.063800
Author contributions: A.A., F.M., A.P., A.J.R.H. and M.M. designed research. A.A., F.M., M.L., A.J., C.S. and M.M. performed research. A.A., F.M., A.R.F., F.P., C.S. and M.M. analyzed data. A.J.R.H. and M.M. wrote the paper.

[1] The abbreviations used are: HLA, human leucocyte antigen; CTL, cytotoxic T lymphocytes; PTM, post-translational modification; EThcD, electron transfer/higher-energy collision dissociation; FDR, false discovery rate.

ASBMB MOLECULAR & CELLULAR PROTEOMICS MCP

et al. characterized 11 phosphorylated ligands, eight of them associated to HLA-B*07, from either a renal carcinoma or a lymphoid cell line (4). Finally, Cobbold *et al.* identified 10 and 85 phosphopeptides restricted, respectively, by HLA-A*02 and HLA-B*35 from primary leukemia cells and normal tissue (7).

Here, following up on our previous study on the phospholigandome of HLA-B*40 (8), we combined a phosphopeptide enrichment strategy with high-resolution LC-MS/MS analysis employing the relatively new Electron Transfer/Higher-Energy Collision Dissociation (EThcD) fragmentation scheme (9) to expand the known repertoire of B*40-bound phosphopeptides and to investigate the presentation of phosphorylated ligands by HLA-B*39, -B*27 and -B*07. We earlier showed that EThcD is especially suited for the characterization of the peptidomes bound to HLA-I (10) and HLA-II molecules (11) and their PTMs (12). Using this workflow, we were able to generate a resource of over 260 phosphorylated HLA-I ligands.

These sets of phosphopeptides showed remarkable similarities, displaying a substantial enrichment of phosphorylation at P4 and a high frequency of basic residues at P1. Moreover, the phospholigands reported in previous studies involving HLA-A*02 (3) and HLA-B*07 (7) share the same generic features. We further used biochemical binding assays and X-ray crystallography to investigate the interaction of the ligands phosphorylated at P4 with the HLA-B*40 binding groove. Overall, our data suggest a common structural mechanism to explain the molecular features observed in all the HLA-B-bound phosphopeptidomes studied so far, and provide us with new rules for predicting phosphorylated HLA epitopes.

EXPERIMENTAL PROCEDURES

Cell Lines and Monoclonal Antibodies—HMy2.C1R (C1R) is a human lymphoid cell line with no expression of its endogenous HLA class I genes except HLA-C*04:01 and HLA-B*35:03, which are expressed at very low levels (13). The stable C1R transfectants expressing HLA-B*40:02 (C1R-B*40) and HLA-B*39:01 (C1R-B*39) have been described previously (8, 14). Cells were grown in DMEM medium supplemented with 7.5% FCS (both from Sigma-Aldrich, St Louis, MO) at 37 °C in roller bottles up to a density of 1.5×10^6 cells/ml. The human lymphoblastoid cell line GR (HLA-A*01, -A*03, -B*07, -B*27, -C*02, -C*07) was grown in RPMI 1640 medium to a total number of 10^9 cells. The monoclonal antibody W6/32 (IgG2a specific for a monomorphic HLA class I determinant) has been described elsewhere (15) and was coupled to CNBr-activated Sepharose beads (GE healthcare, Chicago, IL) following manufacturer's instructions.

Experimental Design and Statistical Rationale—The experimental workflow employed was designed to identify with very high confidence as many HLA-B-bound phosphopeptides as possible. In a first set of experiments, we isolated the peptidome associated to HLA-B*40 from the C1R-B*40 cell line. Three independent biological replicates were carried out. Each of the three independently isolated peptide pools were subjected to phosphopeptide enrichment and the enriched and nonenriched fractions were analyzed by LC-MS/MS using EThcD as fragmentation mode. The results from the MS/MS ion search were filtered at a FDR ≤ 1% at the peptide level. In a second set of analyses, we conducted a single-replicate analysis of the peptidome and phosphopeptidome bound to HLA-B*39 in the C1R-B*39 cell line. Enrichment of phosphopeptides was carried out as described for HLA-B*40. In this case, CID was used as fragmentation scheme. Because CID is not particularly well suited for the identification of phosphorylated peptides, candidate spectra were inspected manually and those that seemed to correspond to *bona fide* phosphopeptides were confirmed or rejected by fragmentation of the corresponding synthetic peptide. Finally, we isolated and analyzed by LC-MS/MS the HLA-I peptidome of the GR cell line. This was done with a single biological replicate that was fractionated by strong cation exchange chromatography before LC-MS/MS analysis. EThcD fragmentation was used for peptide fragmentation and results were filtered at a FDR ≤ 1% as described for HLA-B*40.

Isolation of the HLA Class I-Bound Peptide Pools—The immunoprecipitation of HLA class I molecules from C1R-B*39 and C1R-B*40 cells and the isolation of their associated peptidomes was performed as described elsewhere (8). Briefly, about 10^{10} cells were lysed in 20 mM TRIS, 150 mM NaCl (pH 7.5) containing 1% Igepal CA-630 (Sigma-Aldrich) and a mixture of phosphatase (PhosSTOP; Roche, Basel, Switzerland) and protease (Complete; Roche) inhibitors. The lysate was then subjected to differential centrifugation for 10 min at 2000 × *g*, 30 min at 10,000 × *g*, and 1 h at 100,000 × *g*. Afterward, the supernatant was incubated first with TRIS-blocked Sepharose and then with Sepharose beads coated with the monoclonal antibody W6/32. HLA-I-bound peptides were eluted with 0.1% aqueous TFA and filtered through Vivacon 2 devices (Sartorius Stedim, Goettingen, Germany). Peptides were then concentrated in a speedvac, desalted with an OMIX C18 tip (Varian, Palo Alto, CA), and dried to completeness. Phosphopeptide enrichment was conducted as previously described (8). In brief, the peptide pools eluted from HLA-B*40:02 or HLA-B*39:01 were subjected sequentially to both IMAC (Fe^{3+}) and TiO_2 enrichment. Three fractions corresponding to the IMAC eluate, TiO_2 eluate and flow-through were collected, desalted, speedvac dried, and dissolved in 0.1% formic acid before LC-MS analysis. Three independent biological replicates were carried out with the C1R-B*40 cell line and a single one with C1R-B*39.

The isolation of the HLA-I-bound peptidome from the GR cell line was carried out as previously described (10). In brief, cells were lysed for 60 min in Pierce IP lysis buffer (Thermo Scientific, Waltham, MA) and the lysate was cleared by centrifugation for 10 min at 12,800 × *g* and for 60 min at 20,000 × *g* and the supernatant was passed through a 0.45 μm filter. The sample was precleared by sequential incubation with TRIS-blocked Sepharose beads and beads coupled to normal mouse serum. The flow-through was then incubated with Sepharose beads coupled to the monoclonal antibody W6/32. The HLA class I complexes were eluted from the beads with 10% acetic acid and peptides were filtered through a 10 kDa cut-off membrane. Afterward, the peptide pool was fractionated using strong cation exchange chromatography using a Hypercarb trapping column (5 × 0.2 mm, 7 μm particle size; Thermo Scientific) and a SCX column (12 × 0.02 cm, polysulfoethyl aspartamide, 5 μm particle size; Poly LC, Columbia, MD) before analysis by LC-MS/MS.

Cell Lysis, Trypsin Digestion and Phosphopeptide Enrichment for Full Proteome and Phosphoproteome Analysis—To generate data of the full proteome and phosphoproteome of the C1R-B*40 cell line, we performed peptide-centric LC-MS/MS analysis. Therefore, C1R-B*40 cells were re-suspended in lysis buffer (8 M urea, 50 mM ammonium bicarbonate, pH 8.0) supplemented with a mixture of phosphatase (PhosStop, Roche) and protease (Complete mini EDTA free, Roche) inhibitors. The lysate was sonicated three times on ice and centrifuged at 20,000 × *g* at 4 °C for 20 min. The soluble fraction was collected and the protein concentration was determined by Bradford assay. Proteins were reduced with 2 mM DTT at 56 °C for 25 min and alkylated with 4 mM iodacetamide at room temperature for 30 min in

the dark. A 1:75 ratio of Lys-C was used for a digestion at 37 °C for 4h. Afterward, the sample was diluted four times in 50 mM ammonium bicarbonate to reduce urea concentration below 2 M. Then, trypsin was added at ratio of 1:100 and digestion proceeded overnight at 37 °C. Finally, the reaction was quenched with formic acid at a final concentration of 1%. Aliquots of about 200 μg were desalted using C18 Sep-Pak cartridges (Waters, Milford, MA), dried in a speedvac, and stored at −20 °C until further processing. The Ti^{4+}-IMAC phosphopeptide enrichment was executed by following the procedure established by Zhou *et al.* (16).

LC-MS/MS Analysis Using EThcD—The 3 biological replicates of each fraction (Flow-through, IMAC eluate, and TiO_2 eluate) of the B*40-associated peptide pool were analyzed directly by nanoscale LC-MS/MS using a Thermo Scientific EASY-nLC 1000 (Thermo Scientific) coupled to an ETD- and EThcD-enabled LTQ Orbitrap Elite mass spectrometer (Thermo Scientific) using a 20 × 0.1 mm trapping column (Reprosil C18, 3 μm; Dr. Maisch, Ammerbuch-Entringen, Germany) and a 50 × 0.005 cm analytical column (Poroshell-C18, 2.7 μm; Agilent, Santa Clara, CA). Solvent A and B were 0.1% formic acid in water and acetonitrile, respectively. For the analysis of the IMAC and TiO_2 fractions a gradient of 2 to 27% solvent B was used with a total run time of 2 h. For the analysis of the more complex flow through fractions a gradient of 0 to 27% solvent B in 160 min was employed with a total run time of 3 h. Full MS spectra (*m/z* 300–1200) were acquired in the Orbitrap analyzer at a resolution of 60,000 (FWHM). The 10 most abundant precursor ions—excluding unknown and 1+ charge states—were selected for EThcD fragmentation. The maximum ion accumulation time for MS scans was set to 200 ms and for MS/MS scans to 1500 ms. Fragment ions were detected in the Orbitrap at a resolution of 15,000 (FWHM). Dynamic exclusion was enabled with a repeat count of 1 and duration of 60 s.

The tryptic peptides and phosphopeptides derived from the proteome of the C1R-B*40 cell line were analyzed in a Thermo Scientific EASY-nLC 1000 (Thermo Scientific) coupled online to an ETD-enabled Orbitrap Fusion mass spectrometer employing the same columns and solvents described above. Peptides were fractionated using the following gradient: 2 to 3% solvent B in 8 min, 3 to 8% solvent B in 50 min, 8 to 12% solvent B in 20 min, 12 to 15% solvent B in 10 min, 15 to 19% solvent B in 10 min, 19 to 24% solvent B in 10 min, and 24 to 30% solvent B in 10 min. Total analysis time was 140 min. Full MS spectra (*m/z* 350–1500) were acquired in the Orbitrap analyzer at a resolution of 60,000 (FWHM). A top speed method was used, where the most abundant precursor ions were selected within a 5 s duty cycle for the phosphopeptide-enriched samples and 3 s for the full proteome. Precursors were isolated in the quadrupole with a window of 1.6 Da. The phosphopeptide enriched fractions were analyzed with two different methods, either only EThcD with supplemental activation (40% HCD supplemental collision energy) with an Orbitrap readout at resolution of 15 000 (FWHM) or a decision tree procedure carried out similarly to Frese *et al.* (17) with the instrument switching between HCD and ETciD (supplemental activation 15%) and, in both cases, an ion trap readout. For both EThcD and ETciD fragmentations, the ETD reaction time was calibrated based on the precursor charge. The maximum ion accumulation time for MS scans was set to 50 ms and for MS/MS scans to 35 ms. AGC targets were set to 400,000 for the MS and 10,000 for the MS/MS. The maximum injection time for the MS/MS scan was 70 ms and 100 ms for the decision tree and the EThcD setup, respectively. When fragment ions were detected in the linear ion trap the rapid scan rate was used. For the nonenriched sample the full MS parameters were the same as described above but only HCD fragmentation was used with 35% collision energy and with the MS/MS readout in the ion trap. In this case the MS/MS AGC target was set to 30,000 whereas the maximum injection time was 35 ms.

LC-MS/MS Analysis Using CID—The HLA-B*39:01-associated peptidome and phosphopeptidome were analyzed in a nano-LC Ultra HPLC (Eksigent, Framingham, MA) coupled online with a 5600 triple TOF mass spectrometer (AB Sciex, Framingham, MA) using a C18 chromXP trapping column (350 μm × 0.5 mm, 3 μm, Eksigent) and a C18 chromXP column (75 μm × 150 mm, 3 μm, Eksigent). Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. Peptides were fractionated at a flow-rate of 300 nl/min at 40 °C. A 50% of each sample was analyzed under gradient elution conditions consisting of 2% B for 1 min, a linear increase to 30% B in 109 min, a linear increase to 40% B in 10 min, a linear increase to 90% B in 5 min and 90% B for 5 min. Each acquisition cycle included a survey scan of 250 ms between 350 and 1250 m/z units and a maximum of 50 MS2 spectra scanning between 100 and 1500 *m/z* units. The remaining half-samples were fractionated using the following gradient: 2% B for 1 min, a linear increase to 30% B in 181 min, a linear increase to 40% B in 23 min, a linear increase to 90% B in 15 min, and 90% B for 10 min. Under these chromatographic conditions, the mass spectrometer acquired an MS scan (350–1250 *m/z*) of 250 ms and a maximum of 25 MS/MS scans (100–1500 *m/z*) of 100 ms per cycle. The HPLC and the mass spectrometer were respectively controlled with the Eksigent Control (version 3.12, Eksigent) and the Analyst TF software (version 1.7, Eksigent).

MS/MS Ion Search and Peptide Identification—Peptide identification was carried out as previously described (18). Raw MS/MS data were converted to mgf files with Peakview 1.2 (AB SCiex) and searched against a concatenated target-decoy database containing the 88,266 Uniprot entries of the *Homo sapiens* complete proteome set (as of November 5, 2013) and their corresponding reverse sequences. Five independent search engines were used: MASCOT (Matrix Science, version 2.4.0), OMSSA (National Center for Biotechnology Information, version 2.1.9) (19), X!Tandem2 (The Global Proteome Machine Organization, version 2013–02-01–1) (20), X!Tandem2 with k-score plugin (LabKey Software, version 2.3–7806) (21), and Myrimatch (Mass Spectrometry research Center, Vanderbilt University, version 2.2.140) (22). The datasets corresponding to HLA-I-bound peptides were searched without enzyme restriction and with an MS tolerance of 0.01 Da and an MS/MS tolerance of 0.02 Da. The following variable modifications were considered: oxidation of methionine, protein N-terminal acetylation, pyroglutamic acid formation from N-terminal glutamine or glutamic acid and phosphorylation of serine, threonine and tyrosine. For the analysis of the phosphoproteome of C1R-B*40 cells, trypsin was selected as enzyme allowing up to 2 missed cleavages. The variable modifications specified were: oxidation of methionine, protein N-terminal acetylation, pyroglutamic acid formation from N-terminal glutamine or glutamic acid and phosphorylation of serine, threonine, and tyrosine. MS tolerance was set to 10 ppm. MS/MS tolerance was set to 0.05 Da (EThcD) or 0.6 Da (ETciD and HCD). The individual outputs of the search engines were combined by converting each engine-specific scoring scheme to a common probability-based scale as previously described (23). The *p* values of the peptide-spectrum matches were computed based on score distribution models and identifications were filtered at a FDR ≤ 1% at the peptide level.

For the identification of HLA-B*39-bound phosphopeptides, all the spectra matching phosphorylated sequences with His or Arg at P2 were manually inspected. Those that showed signals that could derive from the neutral loss of the phosphate group were selected and compared with the MS2 spectra of the corresponding synthetic peptides. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (24) with the data set identifier PXD005084. The annotated MS2 spectra of all the

ASBMB MOLECULAR & CELLULAR PROTEOMICS MCP

phosphopeptides reported have been uploaded to the Zenodo repository (https://zenodo.org) with the identifier zenodo.166009.

*In Silico Prediction of HLA-B*40 Binders*—The phosphorylated tryptic peptides identified from the proteome of C1R-B*40 cells were mapped in the sequences of their parental proteins (extracted from Uniprot) and a window of nine residues was scanned around the pSer or pThr residue, so that it occupied the nine positions of the window. A hit was considered positive if the corresponding sequence matched the following pattern (in PROSITE format): x-E-x(6)-[FLIVM]. The position of pSer or pThr was recorded, and the corresponding distribution of positions for the whole set of 10,154 phosphorylated peptides calculated. A similar procedure was repeated for windows of length 10 and 11 (*i.e.* patterns x-E-x(7)-[FLIVM] and x-E-x(8)-[FLIVM], respectively).

Peptide Synthesis—Peptides were synthesized on an automated Multipep peptide synthesizer (Intavis, Koeln, Germany) using standard F-moc chemistry. Afterward, they were purified by reversed phase chromatography in a Smartline HPLC (Knauer, Berlin, Germany) equipped with a 218TP52 C18 column (Vydac, Deerfield, IL). Peptides intended for binding assays were quantified by amino acid analysis in a Biochrom 30 analyzer (Biochrom, Cambridge, UK). The peptide GEFGGCGSV was labeled after synthesis with 5-iodoacetamidofluoresceine (Thermo Scientific) following the manufacturer's instructions and quantified by absorbance at 491 nm as previously described (25).

Binding Assays—The peptide binding assays were carried out exactly as described elsewhere (8). After acid stripping of the C1R-B*40 cells, 2 μg/ml beta-2-microglobulin (β2m; Merck Millipore, Darmstadt, Germany), different concentrations of the test peptides and 400 nM of a fluorescent B*40 binder (GEFGGXGSV, X = fluorescein labeled cysteine) were added and incubated with the cells overnight at 4 °C. After two washes with PBS, fluorescence was measured in a Cytomics FC500 flow cytometer (Beckman Coulter, Brea, CA). Inhibition curves were plotted and IC_{50} values (the concentration that yields 50% inhibition of the binding of reference peptide) were estimated as described by Kessler *et al.* (25). The statistical significance of the differences in binding affinity was assessed with a two-tailed unpaired *t* test with Welch's correction.

Recombinant Protein Expression—The DNA coding for residues 1–274 of the HLA-B*40:02 heavy chain was amplified by PCR from a full length cDNA, cloned in the pET-22b vector (Merck Millipore) using its NdeI site and transformed into the *E. coli* strain Rosetta-gami(DE3)pLys. A cDNA encoding β2m cloned in the pET-30a vector (Novagen) was kindly provided by Prof. James McCluskey (University of Melbourne, Parkville, Australia) and was transformed in the *E. coli* strain BL21(DE3)pLys. Bacteria were grown in LB medium supplemented with 50 μg/ml kanamycin (β2m) or 100 μg/ml ampicillin (heavy chain). Protein expression was induced when the cultures reached an OD_{600} = 0.6 by addition of 0.5 mM IPTG. Bacteria were harvested 4 h (β2m) or 5 h (heavy chain) after induction and were stored at −80 °C until further processing.

Bacterial lysis and purification of the inclusion bodies were carried out as described previously (26). Pellets were resuspended in 50 mM TRIS, 0.2 M NaCl, 5 mM EDTA, 5 mM DTT, pH 8.0 and lysozyme was added at a final concentration of 1 mg/ml. After 15 min on ice, Triton X-100 and deoxycholic acid were added at a final concentration of 0.5%. Bacteria were lysed with two freeze-and-thaw cycles and sonicated twice in 1 min cycles. Insoluble material was pelleted down by centrifugation and washed first with 50 mM TRIS, 0.1 M NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton X-100, pH 8.0 then with 50 mM TRIS, 2 M NaCl, 1 M urea, 1 mM EDTA, 1 mM DTT, pH 8.0 and finally with 50 mM TRIS, 1 mM EDTA, 1 mM DTT, pH 8.0. Four washes were performed with each buffer and—after each centrifugation—the pellets were resuspended and sonicated twice in 1 min cycles. After the last centrifugation step, the insoluble protein pellets were dissolved in 25 mM MES, 8 M urea, 10 mM EDTA, 1 mM DTT, pH 6.5. Protein was quantified by absorbance at 280 nm.

Refolding and Purification—Refolding of the trimeric complex was carried out similarly as described by Reid *et al.* (27) in 1 liter of 100 mM TRIS, 400 mM arginine, 2 mM EDTA, 6.3 mM cysteamine, 3.7 mM cystamine, 0.1 mM PMSF, pH 8.3 with 15 mg of heavy chain, 20 mg of β2m, and 10 mg of peptide (molar ratio 1:4:20). First, 5 mg of β2m were left to refold for 12 h at 4 °C. The heavy chain, the peptide and the remaining β2m were pooled and added to the refolding mixture in 4 equal pulses at 12 h intervals. Afterward, the mixture was concentrated first with 62 mm Omega disks (Pall, Port Washington, NY) to an approximate volume of 50 ml and then with Vivaspin Turbo 15 centrifugal filters (Sartorius Stedim) to a final volume of 3 ml and was fractionated using sephadex G75 or G200 gel filtration columns (GE Healthcare). The fractions corresponding to the properly folded complexes were pooled and concentrated up to 15–20 mg/ml using Vivaspin Turbo 4 devices (Sartorius Stedim).

Crystallization—The crystallization of HLA-B*40 in complex with the peptides REF(pS)KEPEL or REFSKEPEL was performed by vapor diffusion in sitting drop. To that end, 150 nl of the purified trimeric complexes were mixed with 150 nl of every crystallization solution included in the Pi-Minimal kit (Jena Bioscience, Jena, Germany) using a Tecan Genesis RSP 150 robot (Tecan, Männedorf, Switzerland). After 3 days, crystals of both complexes were observed under the following conditions: 8% ethylene glycol, 20% PEG 5000, 150 mM TRIS, pH 8.0 (REF(pS)KEPEL), and 70 mM potassium thiocyanate, 23% PEG 8000, 150 mM TRIS, pH 8.5 (REFSKEPEL). Crystals were collected and frozen in liquid nitrogen using 20% ethylene glycol (REF(pS)KEPEL) or 20% glycerol (REFSKEPEL) as cryo-protector.

X-ray Diffraction and Data Processing—X-ray diffraction experiments were carried out using the BL13-XALOC beam-line of the ALBA synchrotron (28) with a wavelength of 1 Å, an exposure time of 250 ms and increments of the rotation angle of 0.25°. The recorded images were integrated with XDS (29) and scaled with AIMLESS (CCP4 suite) (30) and the resulting data were processed with PHASER (CCP4 suite) (31) to solve the structures by molecular replacement. To that end, the crystal structure of HLA-B*41:04 (PDB: 3NL5)—97.1% identity with the α1, α2, and α3 domains of HLA-B*40:02—was used as template. When a solution was found, the amino acid sequence was manually replaced using COOT and the structure was refined with PHENIX (v1.10–2148) (32).

RESULTS

*In-depth Characterization of the HLA-B*40-bound Peptidome and Phosphopeptidome*—As a follow-up to our previous work (8), we set out to identify as many HLA-B*40 phospholigands as possible. B*40 was immunopurified from C1R cells stably transfected with this allele and, following peptide elution and phosphopeptide enrichment, its associated peptidome and phosphopeptidome were analyzed by LC-MS/MS employing EThcD for peptide fragmentation. Database searching on the peptidome data allowed the identification of 7375 unique peptides at a FDR < 1% (supplemental Data S1). Of them, 70 (1%), 214 (3%) and 6674 (91%) were classified as B*35, C*04 or B*40 ligands, respectively, according to the binding motif reported for these molecules (8, 33–35). Only 417 sequences (6%) could not be confidently assigned to any allotype.

Following analysis of the phosphopeptide-enriched fraction of the B*40 ligandome, a total of 113 phosphorylated sequences with Glu, Asp, or pSer at P2 -the canonical B*40

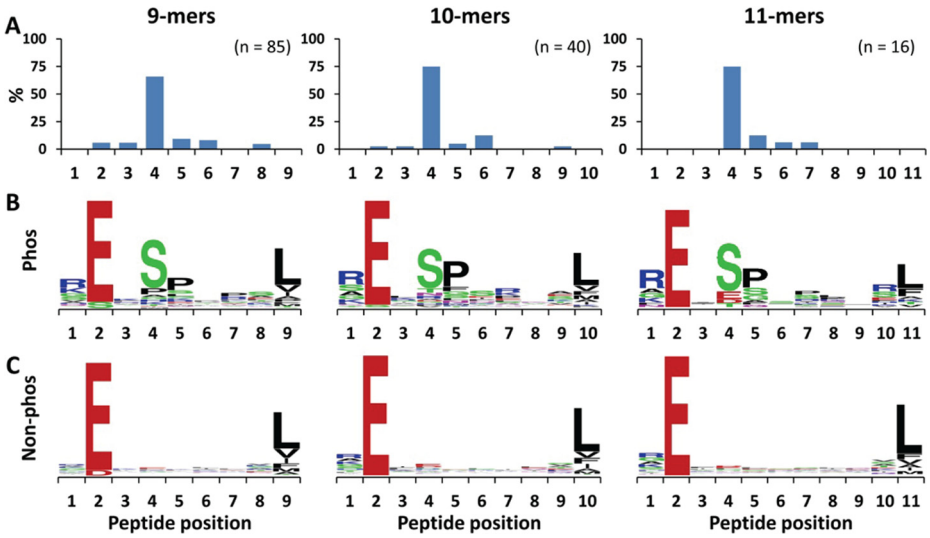

FIG. 1. **Sequence analysis of the phosphorylated and the nonphosphorylated peptides associated to HLA-B*40.** Only the data corresponding to peptides of 9 to 11 residues long are shown. *A*, Frequency distribution of phosphorylation among the identified HLA-B*40 phospholigands. *B* and *C*, Sequence logos of the phosphopeptides (*B*) and the nonmodified peptides (*C*) identified in the HLA-B*40 ligandome.

binding motif (8)– could be identified. To evaluate the reliability of these identifications, we synthesized 26 phosphopeptides and subjected those to analysis by EThcD. The spectra of the B*40-bound phosphopeptides and their synthetic counterparts showed excellent correlation allowing the unambiguous confirmation of at least 25 out of 26 (96%) sequences (supplemental Data S2). We extended the catalogue of the here-identified B*40-associated phosphopeptides, including those from a previously published dataset of 85 phosphorylated ligands identified by us using CID-MS (8). The final compilation comprises 148 unique phosphopeptides derived from 136 proteins (supplemental Data S3).

Next, we queried this list of sequences for specific molecular characteristics. Several remarkable observations emerged when the position of the phosphorylated residue was considered. First, phosphorylation was found at P4 in 101 ligands (68%, Fig. 1*A*). Of them, 65 (64%) carried Pro at P5 and, in the remaining 37 sequences, P5 was enriched for hydrophobic amino acids (*i.e.* Leu, Phe, Val, Ile, and Met) (Fig. 1*B*). Moreover, 56 (55%) of the HLA-B*40 ligands phosphorylated at P4 had also a basic residue, primarily Arg, at P1. Performing a similar analysis on the full nonmodified ligandome it became quite apparent that all these sequence features were unique for the HLA-B*40-bound phosphopeptides (Fig. 1*C*). The sequence motif exhibited by the nonmodified peptides was primarily defined by the two main anchor positions of HLA-B*40, (*i.e.* Glu or Asp at P2 and Met, Phe or aliphatic residues at the C terminus). Beside the predominance of phosphorylation at P4, we also detected, in agreement with our previous report (8), 7 ligands harboring phosphorylation at P2 instead of Glu or Asp, hinting at the fact that pSer could mimic and replace these residues at this major anchor position.

*The Phosphopeptidome Displayed by HLA-B*40 in C1R Cells is Enriched for Ligands Derived from Proteins Involved in Mitosis and Cell Cycle Regulation*—Next, we considered the origin of the 136 source proteins of the detected B*40-bound phosphopeptides and observed that many of those (around 30) related to cell cycle and in particular mitosis, forming a correlated network when analyzed by STRING (36) (Table I). Because C1R cells derive from an Epstein-Barr-transformed lymphoblastoid cell line (37), we wondered if the presumably aberrant phosphorylation pattern of this biological model could explain this observation. To evaluate this possibility, we analyzed the full phosphoproteome of the C1R-B*40 cells. In this comprehensive study, we detected a total of 10,153 unique phosphopeptides (FDR $\leq$ 1%) corresponding to 2369 proteins (supplemental Data S4). When this set of proteins was used as background for the STRING analysis, no specific enrichment of GO terms related to mitosis and cell cycle could be observed (Table I), indicating that the overrepresentation of phosphorylated ligands derived from this sort of proteins is likely a consequence of their high abundance in the phosphoproteome of C1R-B*40 cells.

*Influence of the Cell Kinome in the Shaping of the B*40-bound Phosphopeptidome*—Then, to test if the bias for phosphorylated residues at P4 was the result of a kinase specific recognition motif, we used the tryptic phosphopeptides identified from the C1R-B*40 phosphoproteome to predict *in silico* 2017 sequences of 9 to 11 amino acids that matched the canonical B*40 binding motif (*i.e.* Glu at P2 and Leu, Ile, Val, Met, or Phe at the C terminus). No enrichment of pSer/pThr at P4 or any other position was observed among the predicted binders (Fig. 2). Moreover, when the sequences phosphorylated at P4 were considered, only 50 out of 284 (18%) carried basic residues at P1. In contrast, Pro was somewhat overrep-

TABLE I
*GO enrichment analysis of the source proteins of the B*40-bound phosphopeptides*

GO id	Term	Number of genes	FDR (Genome[a])	FDR (Phosphoproteome[b])
GO:0022402	cell cycle process	35	$4.29 \cdot 10^{-13}$	1.00
GO:0007049	cell cycle	37	$8.61 \cdot 10^{-12}$	1.00
GO:1903047	mitotic cell cycle process	28	$2.52 \cdot 10^{-11}$	1.00
GO:0000278	mitotic cell cycle	29	$3.93 \cdot 10^{-11}$	1.00
GO:0051301	cell division	23	$4.4 \cdot 10^{-11}$	1.00
GO:0007067	mitotic nuclear division	17	$1.02 \cdot 10^{-7}$	1.00
GO:0048285	organelle fission	19	$1.08 \cdot 10^{-7}$	1.00
GO:0000280	nuclear division	18	$2.77 \cdot 10^{-7}$	1.00
GO:1902589	single-organism organelle organization	35	$3.07 \cdot 10^{-6}$	1.00
GO:0016043	cellular component organization	55	$6.19 \cdot 10^{-5}$	1.00
GO:0044260	cellular macromolecule metabolic process	68	$9.25 \cdot 10^{-5}$	1.00
GO:0006996	organelle organization	40	$9.5 \cdot 10^{-5}$	1.00
GO:0044772	mitotic cell cycle phase transition	13	$9.5 \cdot 10^{-5}$	1.00
GO:0044770	cell cycle phase transition	13	$9.67 \cdot 10^{-5}$	1.00
GO:0071840	cellular component organization or biogenesis	55	$9.67 \cdot 10^{-5}$	1.00
GO:0043170	macromolecule metabolic process	71	$1.29 \cdot 10^{-4}$	1.00
GO:0051983	regulation of chromosome segregation	7	$1.55 \cdot 10^{-4}$	1.00
GO:0051726	regulation of cell cycle	21	$1.77 \cdot 10^{-4}$	1.00
GO:0090304	nucleic acid metabolic process	48	$2.31 \cdot 10^{-4}$	1.00
GO:0045786	negative regulation of cell cycle	14	$5.03 \cdot 10^{-4}$	1.00
GO:0010564	regulation of cell cycle process	15	$5.52 \cdot 10^{-4}$	1.00
GO:0008152	metabolic process	82	$6.71 \cdot 10^{-4}$	1.00
GO:0019222	regulation of metabolic process	62	$9.73 \cdot 10^{-4}$	1.00

[a]Analysis using the human genome as background.
[b]Analysis using the C1R-B*40 phosphoproteome as background.

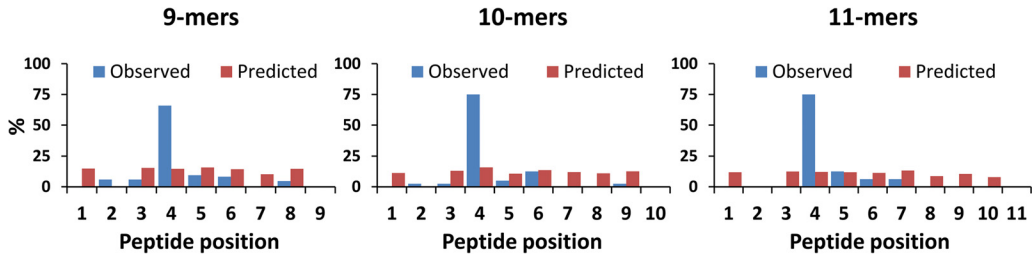

FIG. 2. **Frequency distribution of phosphorylation among the identified HLA-B*40 phosphopeptides (*blue*) and the *in silico* predicted HLA-B*40 binders (*red*).**

resented at P5 (67 sequences, 24%). This suggests that, although the Arg1/Lys1 motif is probably due to a structural constraint imposed by the B*40 binding groove the high frequency of Pro5 most likely reflects the activity of specific kinases.

*Effects of Phosphorylation and Arg at P1 on the Peptide Binding Affinity to HLA-B*40*—A possible explanation for the observed high frequency of phosphorylation at P4 is that the phosphate moiety at this position increases the stability of the HLA-peptide complex. To address this point, seven endogenous phosphorylated ligands and their nonphosphorylated counterparts were synthesized and assayed for binding to HLA-B*40. Most peptides displayed IC_{50} values in the high nanomolar or low micromolar range (0.3–3.8 μM), corresponding to strong and medium affinity binders (Fig. 3*A*). We also identified a weak binder (REASPSRLS) with an IC_{50} of 9.9 μM and 14.6 μM for the nonphosporylated and the phosphorylated species, respectively. In every tested pair of peptides, the IC_{50} of the phosphorylated ligand was slightly higher, revealing that phosphorylation at P4 has a somewhat negative effect on the binding affinity to B*40.

Given that phosphorylation at P4 certainly does not enhance binding, we wondered whether this recurrent feature was the result of a negative selection operating on other peptide positions. Therefore, we evaluated the binding of a set of poly-Gly analogs displaying the B*40 binding motif and a single phosphoserine residue at different positions. As shown in Fig. 3*B*, the highest binding efficiencies were observed when phosphoserine was placed at P4 (0.5 μM) and P8 (0.6 μM) whereas phosphorylation at P5, P6, and P7 led to a moderate decrease of affinity (1.9 μM, 1.8 μM, 1.1 μM, respectively). Finally, phosphorylation at P3 caused a marked increase of the IC_{50} (25.1 μM) reflecting a considerable negative effect on binding.

To assess the effect of a positively charged residue at P1, we tested a second set of poly-Gly analogs harboring Arg at

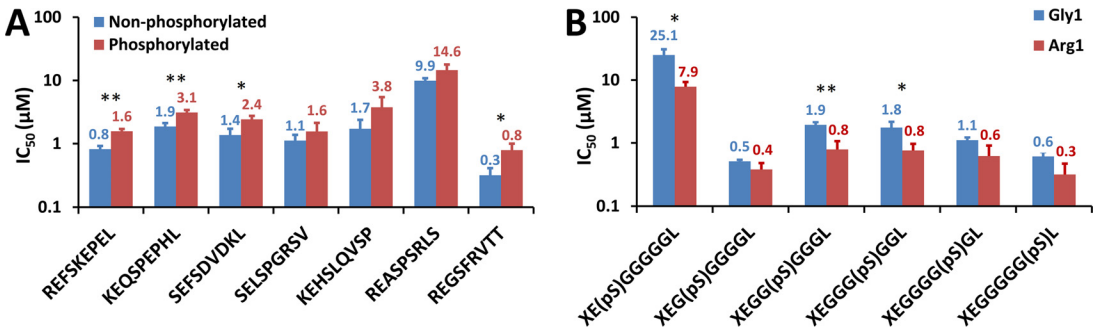

FIG. 3. **Peptide binding assays to HLA-B*40.** *Bars* show the mean ± S.D. of three independent experiments. Statistically significant differences are indicated (** $p < 0.01$; * = $p < 0.05$). *A*, Binding affinities -expressed as IC_{50} values- of seven phospholigands associated to HLA-B*40 and their nonphosphorylated counterparts. In all cases, the pair of peptides exhibit alike affinities although the affinity of the phosphopeptides is somewhat lower. *B*, Binding affinities of 12 phosphorylated poly-Gly analogs with Gly (*blue*) or Arg (*red*) at P1.

the peptide N terminus. Regardless of the phosphorylated position, the presence of this residue at P1 enhanced the binding affinity of each assayed peptide (Fig. 3*B*), partly explaining the increased frequency of this motif in the B*40-associated phosphopeptidome.

*Structures of HLA-B*40 in Complex With an Endogenous Phospholigand or its Nonphosphorylated Counterpart*—We next set out to crystallize HLA-B*40 -a molecule whose crystal structure had not been reported before- in complex with a natural phosphorylated ligand (REF(pS)KEPEL) and its nonphosphorylated counterpart (REFSKEPEL). These ligands were chosen because they displayed the canonical binding motif of B*40 (Glu2 and LeuΩ) and had a high binding affinity according to our assays, facilitating the *in vitro* refolding of the trimeric complex. The two structures were solved and refined at a resolution of 1.5 and 1.8 Å, respectively (Table II). The overall fold of both complexes was almost identical (RMSD = 0.128 Å), with both ligands adopting the typical extended conformation and the peptide N- and C termini anchored in the binding cleft (Fig. 4*A*–4*B*). As expected, P2-Glu and P9-Leu were accommodated in the B and F pockets, respectively. P6-Glu was oriented toward the floor of the binding groove whereas P1-Arg, P4-Ser/P4-pSer, P5-Lys, and P8-Glu were pointed upwards, making these side chains potentially accessible for interaction with the TCR.

Both ligands adopted a very similar conformation, although we observed minor differences in the orientation of residues at P1, P5, and P8. Both P1-Arg and P8-Glu showed two alternative conformations in the phosphorylated peptide (Fig. 4*B*) inducing some changes in the positions of residues Arg62, Glu76 and Glu163 of the heavy chain (Fig. 4*C*). Regarding P5-Lys, the weak electron density around its side chain is probably indicative of a high flexibility. Nevertheless, in the peptide REF(pS)KEPEL, the ε-amino group of this residue tended to be oriented toward P4-pSer, likely influenced by the negative charge of the phosphate group (Fig. 4*B*).

In the phosphorylated ligand, the phosphate moiety of P4-pSer interacted directly with Arg62 and via two ordered water molecules with Arg62 and Glu163 on the heavy chain and P1-Arg on the peptide (Fig. 5*A*–5*B*). In contrast, in the nonphosphorylated peptide, P4-Ser was not involved in any significant interaction with residues of the heavy chain but took part in a network of ordered water molecules, indirectly linking its side chain to that of Arg62 and Glu163 (Fig. 5*C*–5*D*). Finally, in both complexes, the side chain of P1-Arg was stabilized in the A pocket through three main interactions: 1) π-π stacking of its guanidinium group with that of Arg62, 2) hydrophobic interactions of the aliphatic part of its side chain with the indole ring of Trp167 and 3) a salt bridge formed between its guanidinium group and the carboxyl group of Glu163 (Fig. 5). This mode of binding provides at least a partial explanation to the increased affinity for HLA-B*40 of the phosphopeptides with Arg at P1.

Preferred Phosphorylation at P4 is a Landmark of More HLA-B Alleles—Notably, inspecting the scarce and scattered available data from the literature, the strong preference of HLA-B*40 for peptides phosphorylated at P4 can be also observed in other HLA-B allotypes like B*27 (5) or B*07 (7) (Table III). We above showed that Arg62 is a key residue in the stabilization of the phosphate moiety at P4 in HLA-B*40. Because it is conserved in most HLA-B alleles, we hypothesized that this could be the basis of a common phosphorylation motif. To test this hypothesis, we extended our analysis to the phosphopeptidomes associated to HLA-B*39, -B*07, and -B*27, expressed in different cell lines.

First, we isolated the peptidome and phosphopeptidome displayed by HLA-B*39 from C1R cells stably transfected with this allele, as described for HLA-B*40, and analyzed them by LC-MS/MS. A total of 1128 unique peptides of eight to 13 residues long were identified at a FDR ≤ 1% at the peptide level (supplemental Data S5). Of them, 953 (84%) had His or Arg at P2, the canonical binding motif of this allele (14, 38) whereas 14 (1%) and 53 (5%) matched respectively the binding motifs of HLA-B*35 (33) and HLA-C*04 (34, 35). The remaining 108 sequences (10%) could not be confidently assigned to any allotype. Nonetheless, we found 55 peptides (51%) with Gln at P2 in this latter group, suggesting that

TABLE II
Data collection and refinement statistics

	HLA-B*40:02-REF(pS)KEPEL	HLA-B*40:02-REFSKEPEL
PDB ID	5IEH	5IEK
Wavelength (Å)	1.00556	1.00556
Resolution range	24.86–1.5 (1.554–1.5)	25–1.8 (1.864–1.8)
Space group	$P2_12_12_1$	$P2_12_12_1$
UniT-cell	51.044 82.249 110.084 90 90 90	51.26 82.201 111.145 90 90 90
Total reflections	149752 (14732)	88506 (8693)
Unique reflections	74898 (7370)	44288 (4349)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	1.00 (1.0)	1.00 (1.00)
Mean I/sigma(I)	32.48 (8.51)	6.19 (1.63)
Wilson B-factor	13.75	19.74
R-merge	0.01038 (0.07625)	0.04711 (0.2432)
R-meas	0.01468 (0.1078)	0.06662 (0.344)
CC	1 (0.982)	0.996 (0.873)
CC*	1 (0.995)	0.999 (0.965)
Reflections used in refinement	74897 (7370)	44243 (4341)
Reflections used for R-free	3697 (363)	2211 (221)
R-work	0.1611 (0.1791)	0.1613 (0.2208)
R-free	0.1837 (0.2100)	0.1870 (0.2638)
Number of non-hydrogen atoms	3816	3716
Macromolecules	3181	3190
Ligands	4	
Water	621	526
Protein residues	383	385
RMS(bonds)	0.008	0.009
RMS(angles)	0.99	1.04
Ramachandran favored (%)	99	99
Ramachandran allowed (%)	1.3	1
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0.88	3.2
Average B-factor	18.09	24.56
macromolecules	15.77	22.88
ligands	23	
solvent	29.84	34.72

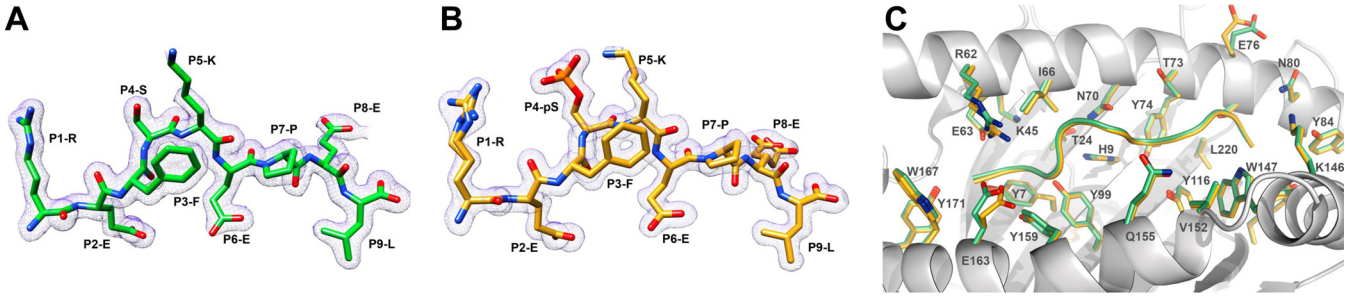

FIG. 4. **Crystal structures of HLA-B*40 in complex with the peptides REF(pS)KEPEL or REFSKEPEL.** *A* and *B*, Structure and electron density map (2Fo-Fc) at 1σ (*blue mesh*) of the peptides REFSKEPEL (*A*), and REF(pS)KEPEL (*B*) anchored to the B*40 binding cleft. *C*, Superposition of the binding grooves of both complexes with the peptides REFSKEPEL (*green*) and REF(pS)KEPEL (*yellow*) displayed in ribbon style.

HLA-B*39 can also bind and present ligands with this suboptimal anchor motif.

The analysis by LC-MS/MS of the phosphopeptide-enriched fractions allowed the characterization of 24 phosphorylated ligands that matched the B*39 binding motif (supplemental Data S5). Because data acquisition was carried out using CID—a fragmentation scheme not particularly well-suited for the identification of phosphopeptides—all these sequences were confirmed by fragmentation of the corresponding synthetic peptide (supplemental Data S6). In this set of ligands, phosphorylation was found at P4 in 18 peptides (75%). Distinct from the B*40-bound phosphopeptidome, only 3 phospholigands (17%) carried positively charged residues at the N terminus. This latter number was, however,

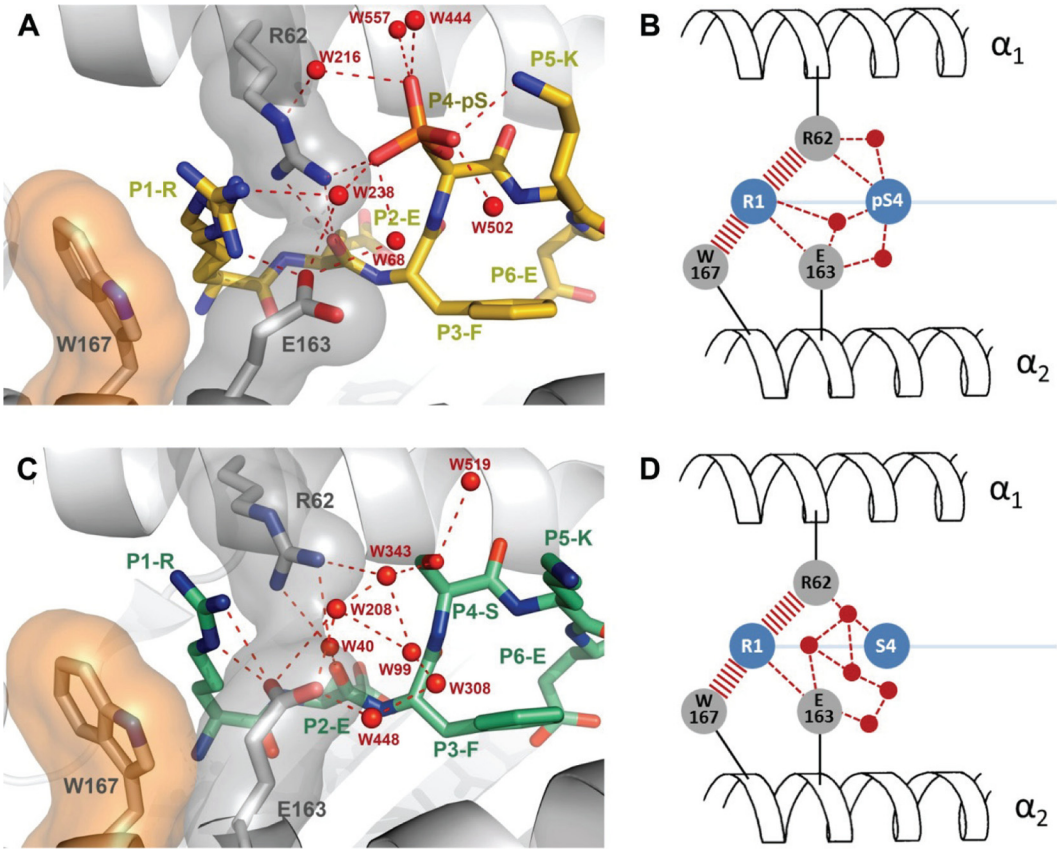

FIG. 5. **Overview of the interactions of residues at P1 and P4 of the peptides REF(pS)KEPEL (*A* and *B*) and REFSKEPEL *C* and *D*, with the binding groove of HLA-B*40.** (A) Interactions involving residues P1-Arg and P4-pSer in the complex B*40-REF(pS)KEPEL. The carbon atoms of the peptide ligand are depicted in *yellow*. *B*, Schematic representation of the interactions displayed in panel *A*. *C*, Interactions involving residues P1-Arg and P4-pSer in the complex B*40-REFSKEPEL. The carbon atoms of the peptide ligand are depicted in *green*. *D*, Schematic representation of the interactions displayed in panel *C*.

TABLE III
Molecular features of the phoshopeptidomes associated to different HLA-B molecules

Allotype	Phosphorylated Ligands	P4 phosphorylation	P4 phosphorylation + basic residue at P1	Preferred motif	Reference
HLA-B*07	6	5 (83%)	5 (100%[a])	-	(5)
	8	5 (63%)	5 (100%[a])	-	(4)
	85	62 (73%)	43 (69%[a])	-	(7)
	32	19 (59%)	13 (68%[a])	(R/K)PRpS	This study
HLA-B*27	6	6 (100%)	5 (83%[a])	-	(5)
	52	37 (65%)	34 (92%[a])	RRXpS	This study
HLA-B*40	148	101 (68%)	56 (55%[a])	REXpS(L/F/M) and XEXpSP	This study
HLA-B*39	24	18 (75%)	3 (17%[a])	X(R/H)XpS	This study

[a]The reported percentage is relative to the number of peptides phosphorylated at P4.

significantly higher than the frequency of basic residues at P1 in the nonphosphorylated ligandome where less than 1% of the peptides had Arg or Lys at this position (Fig. 6).

Finally, we characterized the peptidome and phosphopeptidome displayed by HLA-B*27 and -B*07 in the lymphoblastoid cell line GR. To that end, the HLA-I ligandome of this cell line was fractionated by strong cation exchange chromatography prior to analysis by LC-MS/MS with EThcD fragmentation. A total of 11,041 peptides were identified at a FDR ≤ 1% of which 10,462 (95%) were 8 to 13 residues long. Of them, 4153 (40%) and 2523 (24%) matched the binding motifs of HLA-B*27 (39, 40) and HLA-B*07 (41), respectively (supplemental Data S7). Because the peptide binding motif of HLA-B*27 partially overlaps with that of HLA-C*07 (42), which is also expressed in the GR cell line, it is not possible to assign with absolute confidence a peptide with the mentioned motif to one of these allotypes. Nevertheless, we considered that the vast majority of the phosphopeptides assigned to B*27

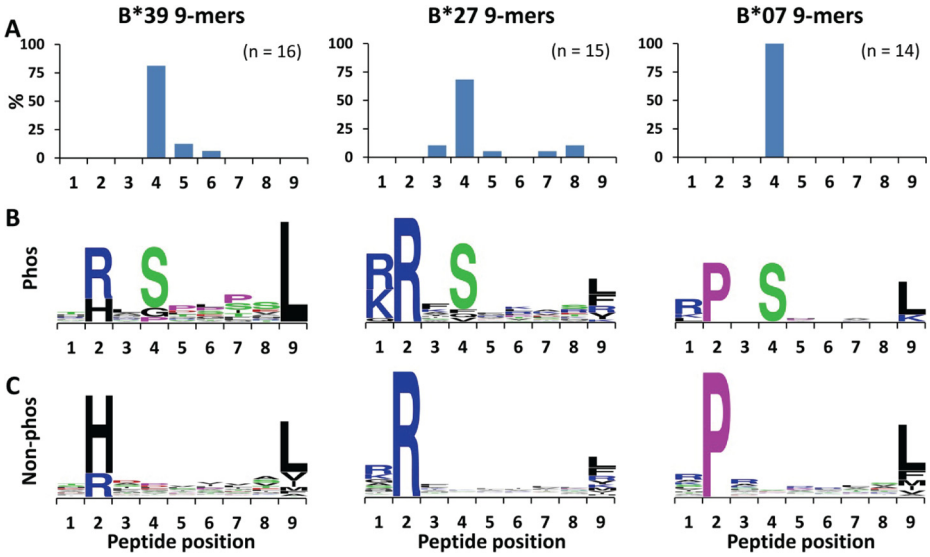

FIG. 6. **Sequence analysis of the phosphorylated and the nonphosphorylated peptides associated to HLA-B*39 (*left column*), HLA-B*27 (*middle column*) or HLA-B*07 (*right column*).** For clarity, only the data corresponding to 9-mer peptides are shown. For other peptide lengths, see supplemental Data S8. *A*, Frequency distribution of phosphorylation at each peptide position in the phospholigandomes bound to these allotypes, revealing the preference for phosphorylation at P4. *B* and *C*, Sequence logos of the phosphopeptides (*B*) and the nonmodified peptides (*C*) identified in the ligandomes of these HLA-B molecules.

are in fact B*27 ligands for two reasons. First, the expression of B alleles is known to be much greater than that of C alleles (43) and thus, the contribution of HLA-C*07 to the immunopeptidome of GR cells should be minor compared with that of B*27. Second, a peptide matching the anchor motifs of two different allotypes is likely to be presented by both of them *in vivo*. Thus, even if some of the phosphopeptides reported are HLA-C*07 ligands, this does not exclude that they are B*27 ligands as well.

Among the B*07 ligands, we could identify 32 phosphorylated peptides (supplemental Data S7) of which 19 (59%) displayed their phosphorylation at P4, with 13 of those (68%) harboring a basic amino acid at the N terminus (Table III and Fig. 6). For HLA-B*27, we were able to identify 52 phospholigands (supplemental Data S7), including 37 (65%) that were phosphorylated at P4 (Table III and Fig. 6). In this case, the preference for basic residues at P1 was even more apparent because 34 (92%) of the peptides phosphorylated at P4 displayed positively charged residues at P1.

All this new data reiterates that the increased phosphorylation at P4 is a general feature of the phosphopeptidomes displayed by different HLA-B allotypes, despite the diversity of their anchoring motifs. Additionally, even though the preference for a basic residue at P1 in the ligands phosphorylated at P4 is common to all the alleles studied so far, the magnitude of this preference is certainly allotype dependent.

DISCUSSION

T-cell-based immunotherapy is an emerging new strategy for cancer treatment, especially in late-stage diseases or those that do not respond to conventional therapies (44). This approach relies critically on the identification of tumor-specific MHC-restricted antigens capable of triggering an effective antitumor response. Neo-antigens arising from tumor-specific mutations are arguably the most obvious targets for these personalized therapies, as they are uniquely associated to tumor development. Nevertheless, because kinase and phosphatase function is often deregulated in tumor cells (45), it has been suggested that phosphorylated antigens could also be ideal candidates for immunotherapy, as their higher abundance may be cancer-related (3, 4). Supporting this hypothesis, Zarling *et al.* showed that $CD8^+$ T-cells directed against two phosphorylated HLA-I epitopes derived from the insulin receptor substrate 2 and the cell cycle regulator CDC25b were able to reduce tumor growth *in vivo* (46).

Yet, the characterization of MHC class I-bound phospholigands remains challenging because of their subtoichiometric levels compared with the nonmodified ligandome and the difficulties associated with their characterization by most common LC-MS/MS workflows (47, 48). In an effort to improve the identification of HLA-I phospholigands, we combined a rather novel peptide fragmentation technique (EThcD), phosphopeptide enrichment and advanced database searching to identify about 150 unique phosphorylated peptides presented by HLA-B*40 on the surface of C1R cells. Reflecting the particularities of the phosphoproteome of this cell line, a significant number of those derived from proteins involved in mitosis and cell cycle control, confirming that alterations in the phosphorylation pattern of proteins relevant to malignant transformation are potentially detectable by T-cells.

ASBMB MOLECULAR & CELLULAR PROTEOMICS MCP

Intriguingly, the phosphopeptidome bound to HLA-B*40 shared some molecular features with those associated to other previously studied MHC class I molecules: phosphorylation occurred mainly at P4 and was frequently accompanied by the presence of a basic residue at P1 (3–5, 7). This prompted us to study the molecular mechanism underlying the presentation of phosphopeptides by this allotype. The binding assays described in our study proved that phosphorylation at P4 has a small negative effect on binding affinity to B*40. Although this result seemingly conflicts with the high occurrence of phosphorylation at this position, the experiments performed with poly-Gly analogs demonstrated that phosphorylation at P4 correlates with higher binding affinity to B*40 when compared with other peptide positions. From a structural point of view, this effect can be explained by the direct contact of the phosphate moiety of P4-Ser with the side chain of Arg62 and its water-mediated interaction with the carboxyl group of Glu163. This mode of stabilizing the phosphate group at P4 shares some similarities with that described for HLA-A*02, where residues Arg65 and/or Lys66 -also localized in the α_1 helix but absent in HLA-B molecules- play a key role in the interaction with the phosphorylated residue (49, 50).

The binding assays performed here also account for the high frequency of basic residues at the N terminus in the B*40-bound phosphopeptidome because the presence of a positively charged residue at this position enhances complex stability, compensating the negative effect on binding caused by phosphorylation at P4. This is probably because of the set of interactions established by the basic residue at P1 with residues of the A pocket. In the two crystal structures presented here, P1-Arg is stabilized through contacts with Arg62, Glu163 and Trp167. Finally, in the phosphorylated ligand, residues P1-Arg and P4-pSer are linked through a water molecule and by a long-distance interaction mediated by residues Arg62 and Glu163, contributing to the overall stability of the complex.

Based on these results, we argued that the conservation of Arg62 in the vast majority of HLA-B allotypes could be the basis for their common preference for peptides phosphorylated at P4. The subsequent analysis of the phosphopeptidomes associated to HLA-B*07, HLA-B*27, and HLA-B*39—which maintain the Arg62 residue—seems to support this hypothesis as they were all found to be enriched in peptides with this molecular feature. Furthermore, the enrichment for Arg or Lys at P1 among the peptides phosphorylated at P4 was also observed in all the HLA-B alleles studied here. Nevertheless, the magnitude of this effect varied greatly among allotypes and correlated with the frequency of basic residues at P1 in the unmodified ligandomes. For instance, in HLA-B*27, the frequency of positively charged residues at P1 among the peptides phosphorylated at P4 was above 90% whereas it was close to 50% in the unmodified peptidome. In contrast, in B*39, only 17% of the ligands phosphorylated at P4 had positively charged residues at the N terminus whereas this frequency was below 1% in the nonphophorylated peptide set. This latter fact might be related to the presence of Thr163 instead of Glu163 in this molecule that would make the A pocket less prone to accommodate positively charged side chains.

By generating the largest data set of HLA-I-associated phosphopeptides so far, from four different allotypes, we could conclude that they share a strong preference for peptides phosphorylated at P4. This may become a key parameter in predicting tumor antigens arising from aberrant phosphorylation. In addition, our data also highlights that each HLA-B allotype has its own binding preferences that may have an effect on the molecular features of the presented phosphopeptides. Therefore, if the peptide binding motif of an MHC-I molecule aligns with the phosphorylation motif of a particular kinase, this allotype may become an indicator of the aberrant activity of this kinase. To illustrate our reasoning, nearly all the phosphorylated peptides presented by HLA-B*27 confirm to the RRXpS motif (Table III), which is a clear hallmark of basophilic kinases such as PKA and PKC (51). In contrast, HLA-B*40 displays a set of phosphorylated ligands that can be separated in the following sequence motifs: REXpS(L/F/M) and XEXpSP. The former resembles the well-known Plk1 kinase motif (52), whereas the latter is informative for proline-directed kinases (53). For B*07 ligands, the prevalent phosphorylation motif is (R/K)PXpS, which loosely confirms to the substrate motif of CDKL5 (54). Therefore, our data seem to predict that individuals expressing specific HLA allotypes may be more prone to present phosphorylated peptides following aberrant function of specific kinases, providing another reason for further research into personalized therapies.

Acknowledgments—We thank Prof. James McCluskey (University of Melbourne, Parkville) for providing the β2m cDNA clone and Dr. Francisco Gavilanes (Universidad Complutense, Madrid) for his expert help in peptide quantification. The X-ray diffraction experiments were performed at XALOC beamline at ALBA Synchrotron with the collaboration of ALBA staff.

* This research was performed within the framework of PRIME-XS, grant number 262067, funded by the European Union 7th Framework Program. The Proteomics Unit of the Spanish National Biotechnology Centre belongs to ProteoRed (PRB^2-ISCIII) and is supported by the FIS grant PT13/0001. FM and AJRH are supported by the project Proteins At Work, a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands (project number 184.032.201). Additionally, AJRH received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 686547 (MSMed) and from the NWO Gravity Program Institute for Chemical Immunology. AA was funded by the JAE-Pre 2011 program of the Spanish National Research Council (CSIC).

S This article contains supplemental material.

¶¶ To whom correspondence may be addressed: E-Mail: A.J.R.Heck@uu.nl. Or, Spanish National Biotechnology Centre (CSIC), 3, Darwin St., Madrid 28049 Spain. Tel.: 34–91-5854540; Fax: E-Mail: mmarcilla@cnb.csic.es.

§§ These authors contributed equally to this work.

The authors declare no conflict of interest.

ASBMB
MOLECULAR & CELLULAR PROTEOMICS
MCP

REFERENCES

1. Robinson, J., Soormally, A. R., Hayhurst, J. D., and Marsh, S. G. (2016) The IPD-IMGT/HLA Database - New developments in reporting HLA variation. *Human Immunol.* **77,** 233–237
2. Engelhard, V. H., Altrich-Vanlith, M., Ostankovitch, M., and Zarling, A. L. (2006) Post-translational modifications of naturally processed MHC-binding epitopes. *Cur. Opin. Immunol.* **18,** 92–97
3. Zarling, A. L., Polefrone, J. M., Evans, A. M., Mikesh, L. M., Shabanowitz, J., Lewis, S. T., Engelhard, V. H., and Hunt, D. F. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14889–14894
4. Meyer, V. S., Drews, O., Gunder, M., Hennenlotter, J., Rammensee, H. G., and Stevanovic, S. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. *J. Proteome Res*. **8,** 3666–3674
5. Zarling, A. L., Ficarro, S. B., White, F. M., Shabanowitz, J., Hunt, D. F., and Engelhard, V. H. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. *J. Exp. Med.* **192,** 1755–1762
6. Andersen, M. H., Bonfill, J. E., Neisig, A., Arsequell, G., Sondergaard, I., Valencia, G., Neefjes, J., Zeuthen, J., Elliott, T., and Haurum, J. S. (1999) Phosphorylated peptides can be transported by TAP molecules, presented by class I MHC molecules, and recognized by phosphopeptide-specific CTL. *J. Immunol.* **163,** 3812–3818
7. Cobbold, M., De La Pena, H., Norris, A., Polefrone, J. M., Qian, J., English, A. M., Cummings, K. L., Penny, S., Turner, J. E., Cottine, J., Abelin, J. G., Malaker, S. A., Zarling, A. L., Huang, H. W., Goodyear, O., Freeman, S. D., Shabanowitz, J., Pratt, G., Craddock, C., Williams, M. E., Hunt, D. F., and Engelhard, V. H. (2013) MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. *Sci. Transl. Med.* **5,** 203ra125
8. Marcilla, M., Alpizar, A., Lombardia, M., Ramos-Fernandez, A., Ramos, M., and Albar, J. P. (2014) Increased diversity of the HLA-B40 ligandome by the presentation of peptides phosphorylated at their main anchor residue. *Mol. Cell. Proteomics* **13,** 462–474
9. Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84,** 9668–9673
10. Mommen, G. P., Frese, C. K., Meiring, H. D., van Gaans-van den Brink, J., de Jong, A. P., van Els, C. A., and Heck, A. J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U.S.A.* **111,** 4507–4512
11. Mommen, G. P., Marino, F., Meiring, H. D., Poelen, M. C., van Gaans-van den Brink, J. A., Mohammed, S., Heck, A. J., and van Els, C. A. (2016) Sampling from the proteome to the HLA-DR ligandome proceeds via high specificity. *Mol. Cell Proteomics* **15,** 1412–1423
12. Marino, F., Bern, M., Mommen, G. P., Leney, A. C., van Gaans-van den Brink, J. A., Bonvin, A. M., Becker, C., van Els, C. A., and Heck, A. J. (2015) Extended O-GlcNAc on HLA Class-I-bound peptides. *J. Am. Chem. Soc.* **137,** 10922–10925
13. Zemmour, J., Little, A. M., Schendel, D. J., and Parham, P. (1992) The HLA-A,B "negative" mutant cell line C1R expresses a novel HLA-B35 allele, which also has a point mutation in the translation initiation codon. *J. Immunol.* **148,** 1941–1948
14. Yague, J., Ramos, M., Vazquez, J., Marina, A., Albar, J. P., and Lopez de Castro, J. A. (1999) The South Amerindian allotype HLA-B*3909 has the largest known similarity in peptide specificity and common natural ligands with HLA-B27. *Tissue Antigens* **53,** 227–236
15. Barnstable, C. J., Bodmer, W. F., Brown, G., Galfre, G., Milstein, C., Williams, A. F., and Ziegler, A. (1978) Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* **14,** 9–20
16. Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A. J., Zou, H., and Mohammed, S. (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **8,** 461–480
17. Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **10,** 2377–2388
18. Marcilla, M., Alvarez, I., Ramos-Fernandez, A., Lombardia, M., Paradela, A., and Albar, J. P. (2016) Comparative analysis of the endogenous peptidomes displayed by HLA-B*27 and Mamu-B*08: Two MHC class I alleles associated with elite control of HIV/SIV infection. *J. Proteome Res.* **15,** 1059–1069
19. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3,** 958–964
20. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
21. MacLean, B., Eng, J. K., Beavis, R. C., and McIntosh, M. (2006) General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* **22,** 2830–2832
22. Tabb, D. L., Fernando, C. G., and Chambers, M. C. (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **6,** 654–661
23. Ramos-Fernandez, A., Paradela, A., Navajas, R., and Albar, J. P. (2008) Generalized method for probability-based peptide and protein identification from tandem mass spectrometry data and sequence database searching. *Mol. Cell. Proteomics* **7,** 1748–1754
24. Vizcaino, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44,** D447–D456
25. Kessler, J. H., Benckhuijsen, W. E., Mutis, T., Melief, C. J., van der Burg, S. H., and Drijfhout, J. W. (2004) Competition-based cellular peptide binding assay for HLA class I. *Current protocols in immunology / edited by John E. Coligan.* [*et al.*] Chapter 18, Unit 18 12
26. Jimenez, D., Roda-Navarro, P., Springer, T. A., and Casasnovas, J. M. (2005) Contribution of N-linked glycans to the conformation and function of intercellular adhesion molecules (ICAMs). *J. Biol. Chem.* **280,** 5854–5861
27. Reid, S. W., Smith, K. J., Jakobsen, B. K., O'Callaghan, C. A., Reyburn, H., Harlos, K., Stuart, D. I., McMichael, A. J., Bell, J. I., and Jones, E. Y. (1996) Production and crystallization of MHC class I B allele single peptide complexes. *FEBS Lett.* **383,** 119–123
28. Juanhuix, J., Gil-Ortiz, F., Cuni, G., Colldelram, C., Nicolas, J., Lidon, J., Boter, E., Ruget, C., Ferrer, S., and Benach, J. (2014) Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* **21,** 679–689
29. Kabsch, W. (2010) *Xds. Acta Crystallogr. D Biol. Crystallogr*. **66,** 125–132
30. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 122–133
31. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
32. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol Crystallogr.* **66,** 486–501
33. Steinle, A., Falk, K., Rotzschke, O., Gnau, V., Stevanovic, S., Jung, G., Schendel, D. J., and Rammensee, H. G. (1996) Motif of HLA-B*3503 peptide ligands. *Immunogenetics* **43,** 105–107
34. Buchsbaum, S., Barnea, E., Dassau, L., Beer, I., Milner, E., and Admon, A. (2003) Large-scale analysis of HLA peptides presented by HLA-Cw4. *Immunogenetics* **55,** 172–176
35. Schittenhelm, R. B., Dudek, N. L., Croft, N. P., Ramarathinam, S. H., and Purcell, A. W. (2014) A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides. *Tissue Antigens* **83,** 174–179
36. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–D452
37. Storkus, W. J., Howell, D. N., Salter, R. D., Dawson, J. R., and Cresswell, P. (1987) NK susceptibility varies inversely with target cell class I HLA antigen expression. *J. Immunol.* **138,** 1657–1659

38. Falk, K., Rotzschke, O., Takiguchi, M., Gnau, V., Stevanovic, S., Jung, G., and Rammensee, H. G. (1995) Peptide motifs of HLA-B38 and B39 molecules. *Immunogenetics* **41,** 162–164
39. Lopez de Castro, J. A., Alvarez, I., Marcilla, M., Paradela, A., Ramos, M., Sesma, L., and Vazquez, M. (2004) HLA-B27: a registry of constitutive peptide ligands. *Tissue Antigens* **63,** 424–445
40. Ben Dror, L., Barnea, E., Beer, I., Mann, M., and Admon, A. (2010) The HLA-B*2705 peptidome. *Arthritis Rheumatism* **62,** 420–429
41. Huczko, E. L., Bodnar, W. M., Benjamin, D., Sakaguchi, K., Zhu, N. Z., Shabanowitz, J., Henderson, R. A., Appella, E., Hunt, D. F., and Engelhard, V. H. (1993) Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. *J. Immunol.* **151,** 2572–2587
42. Rasmussen, M., Harndahl, M., Stryhn, A., Boucherma, R., Nielsen, L. L., Lemonnier, F. A., Nielsen, M., and Buus, S. (2014) Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J. Immunol.* **193,** 4790–4802
43. Apps, R., Meng, Z., Del Prete, G. Q., Lifson, J. D., Zhou, M., and Carrington, M. (2015) Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *J. Immunol.* **194,** 3594–3600
44. Wang, M., Yin, B., Wang, H. Y., and Wang, R. F. (2014) Current advances in T-cell-based cancer immunotherapy. *Immunotherapy* **6,** 1265–1278
45. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* **411,** 355–365
46. Zarling, A. L., Obeng, R. C., Desch, A. N., Pinczewski, J., Cummings, K. L., Deacon, D. H., Conaway, M., Slingluff, C. L., Jr, and Engelhard, V. H. (2014) MHC-restricted phosphopeptides from insulin receptor substrate-2 and CDC25b offer broad-based immunotherapeutic agents for cancer. *Cancer Res.* **74,** 6784–6795
47. Boersema, P. J., Mohammed, S., and Heck, A. J. (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.* **44,** 861–878
48. Larsen, M. R., Trelle, M. B., Thingholm, T. E., and Jensen, O. N. (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* **40,** 790–798
49. Mohammed, F., Cobbold, M., Zarling, A. L., Salim, M., Barrett-Wilt, G. A., Shabanowitz, J., Hunt, D. F., Engelhard, V. H., and Willcox, B. E. (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. *Nat. Immunol.* **9,** 1236–1243
50. Petersen, J., Wurzbacher, S. J., Williamson, N. A., Ramarathinam, S. H., Reid, H. H., Nair, A. K., Zhao, A. Y., Nastovska, R., Rudge, G., Rossjohn, J., and Purcell, A. W. (2009) Phosphorylated self-peptides alter human leukocyte antigen class I-restricted antigen presentation and generate tumor-specific epitopes. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 2776–2781
51. Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S. G., and Pandey, A. (2007) A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* **25,** 285–286
52. Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E., and Nishida, E. (2003) Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* **278,** 25277–25280
53. Weigel, N. L., and Moore, N. L. (2007) Kinases and protein phosphorylation as regulators of steroid hormone action. *Nuclear Receptor Signaling* **5,** e005
54. Katayama, S., Sueyoshi, N., and Kameshita, I. (2015) Critical Determinants of Substrate Recognition by Cyclin-Dependent Kinase-like 5 (CDKL5). *Biochemistry* **54,** 2975–2987

ASBMB MOLECULAR & CELLULAR PROTEOMICS MCP