Characterization of HLA-A*02:01 MHC Immunopeptide Antigens Enhanced by Ultraviolet Photodissociation Mass Spectrometry

Eleanor Watts[1], Gregory K. Potts[2], Damien B. Ready[2], Alayna George Thompson[2], Janice Lee[2], Edwin E. Escobar[1], Melanie J. Patterson[2], and Jennifer Brodbelt[1]

[1]Department of Chemistry, University of Texas at Austin, Austin, TX

[2]AbbVie, Inc. North Chicago, IL

Correspondence to: jbrodbelt@cm.utexas.edu

Supporting Information

Optimization of MS/MS Activation Parameters

Each ion activation method was optimized based on the Byonic scores for each peptide matched, the sequence coverage, and the fragment ion types generated, especially the number of *d* and *w* ions for UVPD and EThcD. The sequence coverages are calculated by Byonic as a part peptide spectral matches (PSM) scoring. This information was retained by running Byonic with debug enabled and then compiled from multiple files into a single table using an R program written in-house. Byonic includes *a, b*, and *y* ions when calculating sequence coverage for UVPD and HCD and *a, b, c, y*, and *z* ions for EThcD. After examining the Byonic scores and sequence coverages for all conditions considered, a handful of top conditions were selected, and the ion types identified in the Byonic output and examined. Only 18 peptides, ones consistently well-matched across all techniques and conditions, were included owing to the time-consuming nature of compiling the ion types (done manually). In the ion type tables, all *d* and *w* ions only represent those associated with diagnostic neutral losses leading to the specific identification of a leucine or isoleucine residue. The highest scoring PSM was used for the ion type figures.

193 nm UVPD conditions were optimized for laser pulse energy and the number of pulses, HCD for the normalized collision energy (%NCE), and EThcD for the activation time. **Figure S1** summarizes the results for the optimization of the laser pulse energy and the number of pulses for 193 nm UVPD. While there were no significant improvements in the sequence coverages above one pulse with a 2 mJ laser energy, the Byonic scores continued to increase as the number of pulses and laser energy increased. Ordinarily, the lack of increase in sequence coverage would be taken as an indicator that no additional informative fragment ions are being identified, and the continued increase in score is explained by an increase in the fragment ion abundances. In this case however, the *d* and *w* ions were of particular interest due to the goal of localizing leucine and isoleucine residues. By examining **Figure S1C**, it is observed that while the number of *a, b*, and *y* ion types that contribute to the Byonic calculation of sequence coverage did not increase significantly as the laser energy and the number of pulses increased, the number of *c, x*, and *z* ions, as well as the diagnostic *d* and *w* ions, continued to increase. While these particular fragment ions do not contribute to the sequence coverage or to the Byonic score, they are produced from higher energy fragmentation pathways, and it follows that higher laser energies and pulse counts might be necessary to access them. For this reason, four pulses 3 mJ were selected for UVPD. While the identification of these additional *d/w* fragment ions was critical for the differentiation of leucine and isoleucine, it may not be sufficient reason to add these additional ion types to the Byonic scoring algorithm. Searches based on the *a, b*, and *y* types yielded the largest number of matches and therefore seem an appropriate choice when trying to generate an overall score for peptide correctness that balances identifying as many PSMs as possible without increasing the FDR by considering too many ion types of low frequency.

For HCD, examination of the Byonic scores and sequence coverage led to the adoption of 21% NCE. As illustrated in **Figure S2**, a dramatic increase in sequence coverage was observed when NCE was increased from 18% to 21%, and both the Byonic scores and sequence coverages decreased with higher NCE. While 21% NCE is low compared to the NCE often used for immunopeptide analysis, the specific NCE value can vary significantly depending on the instrument used. Previous work done using a Thermo Orbitrap™ Fusion Lumos Tribrid mass spectrometer similar to the one utilized here reported that an NCE as low as 17% can be optimal when analyzing HLA-A*02:01 immunopeptides.[1] **Figure S3** includes the optimization of ETD reaction time for EThcD, segregated based on peptides for which the highest-scoring PSM was either from the 2+ or 3+ charge state. This charge segregation was done because the high charge state dependence of EThcD is well established. Overall, for a range of activation times, relatively modest

changes were observed in both the Byonic scores and sequence coverages. There were no dramatic differences between the findings for the 2+ and 3+ charge states, and a small decrease in score was observed for both charge states when increasing the reaction time from 40 ms to 50 ms. When examining the fragment ions identified in **Figure S3C**, slightly more diagnostic *w* ions were identified using the 40 ms activation time. For this reason, the 40 ms activation time was selected for EThcD."

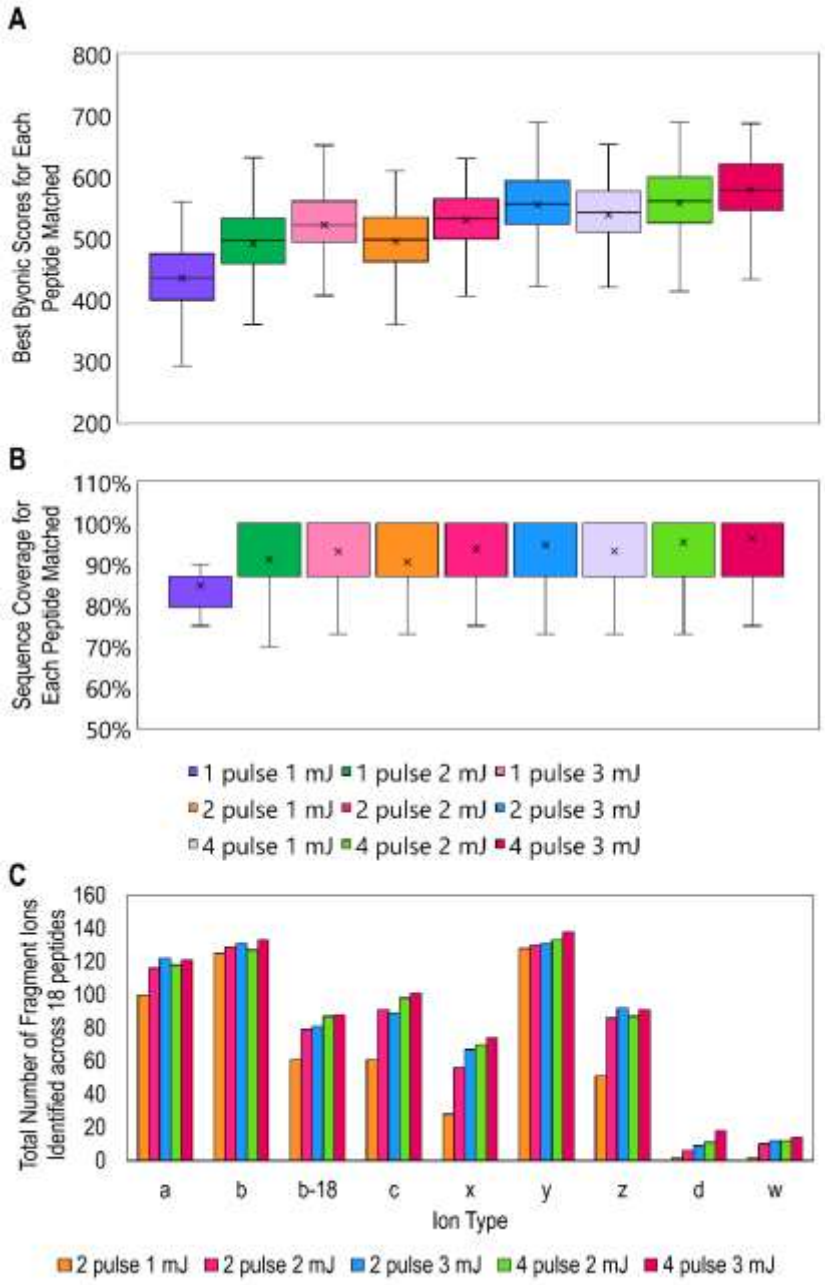


Figure S1: Optimization of laser energy and the number of pulses for 193 nm UVPD, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.

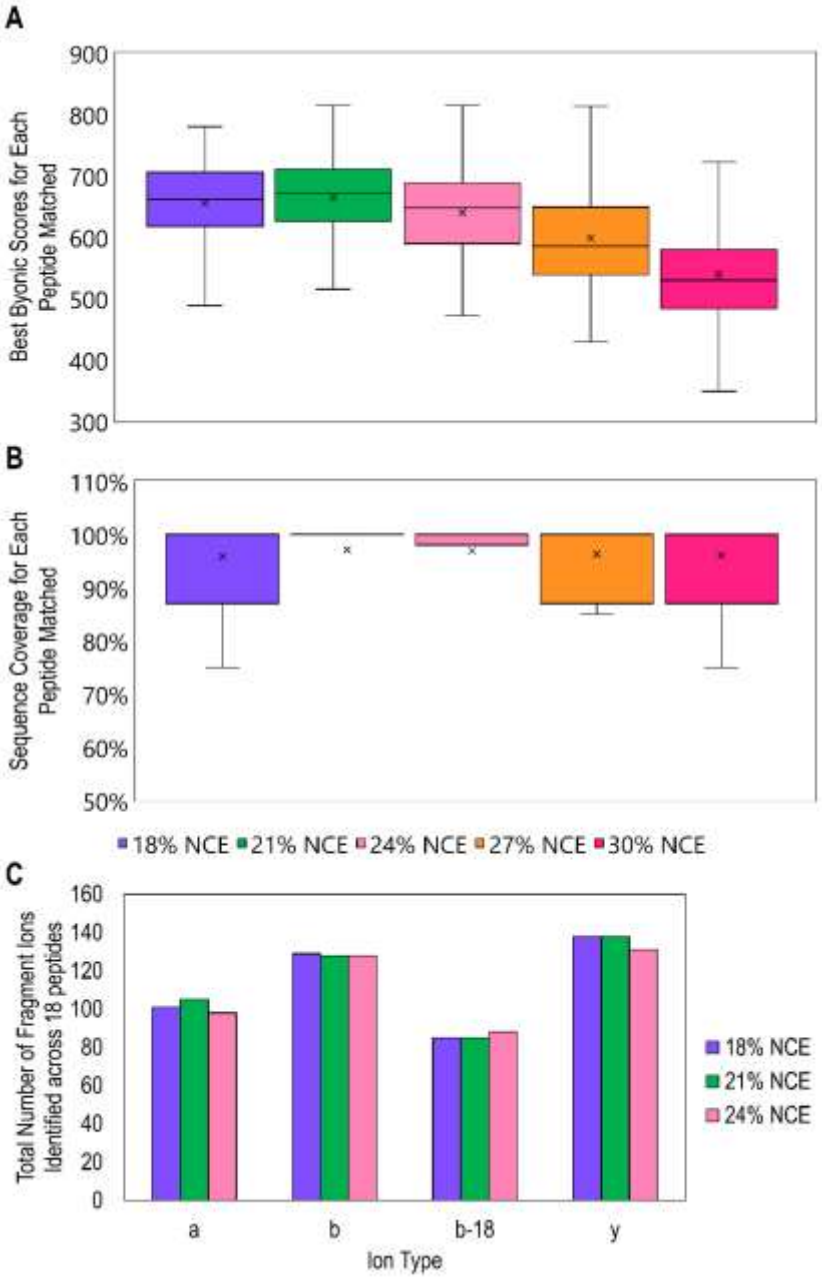


Figure S2: Optimization of normalized collisional energy (% NCE) for HCD, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.

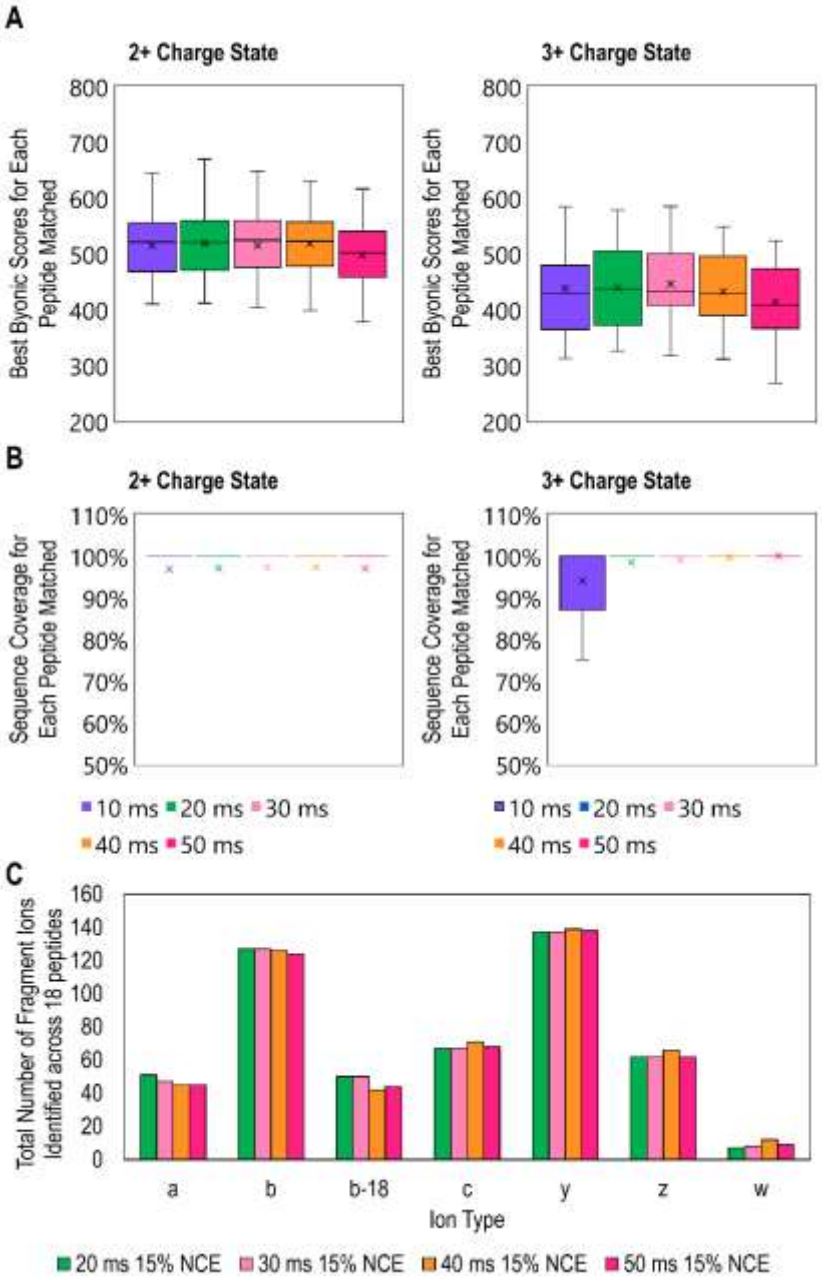
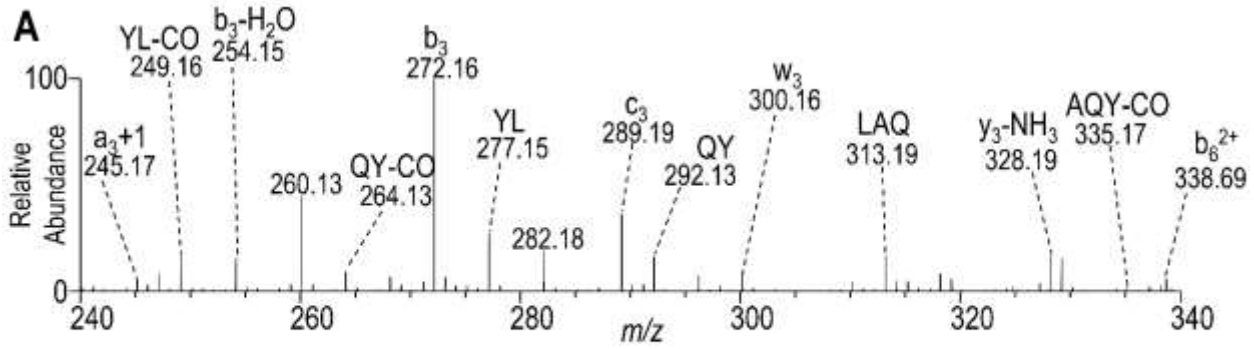


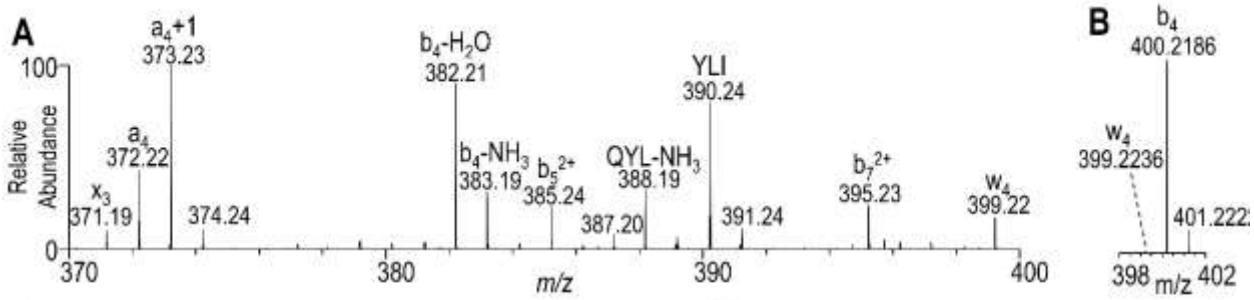
Figure S3: Optimization of activation time for EThcD, broken down by charge state, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Supplemental activation of 15% NCE was applied with all activation times. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.



B

	Theoretical m/z	Observed m/z	Error (ppm)
a_3+1	245.1734	245.1732	0.82
YL-CO	249.1598	249.1594	1.61
b_3-H_2O	254.1499	254.1497	0.79
QY-CO	264.1343	264.1340	1.14
b_3	272.1605	272.1602	1.10
YL	277.1547	277.1544	1.08
c_3	289.1870	289.1867	1.04
QY	292.1292	292.1290	0.68
w_3	300.1554	300.1551	1.00
LAQ	313.1870	313.1867	0.96
y_3-NH_3	328.1867	328.1862	1.52
AQY-CO	335.1714	335.1709	1.49
b_6^{2+}	338.6872	338.6866	1.62

Figure S4: (A) Expansion of region around w_3 fragment ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (2+ charge state) from the BB7.2 immunoprecipitation. (B) Table listing the theoretical *m/z*, observed *m/z*, and ppm error for each identified ion in the spectrum. All theoretical *m/z* values were calculated using Protein Prospector. The mass of the w_3 ion that confirms the identification of isoleucine was calculated by subtracting 29.0391 Da from the mass of the z_3 ion, and the mass of the a_3*+1* ion was calculated by adding 1.0078 Da to the mass of the a_3 ion.



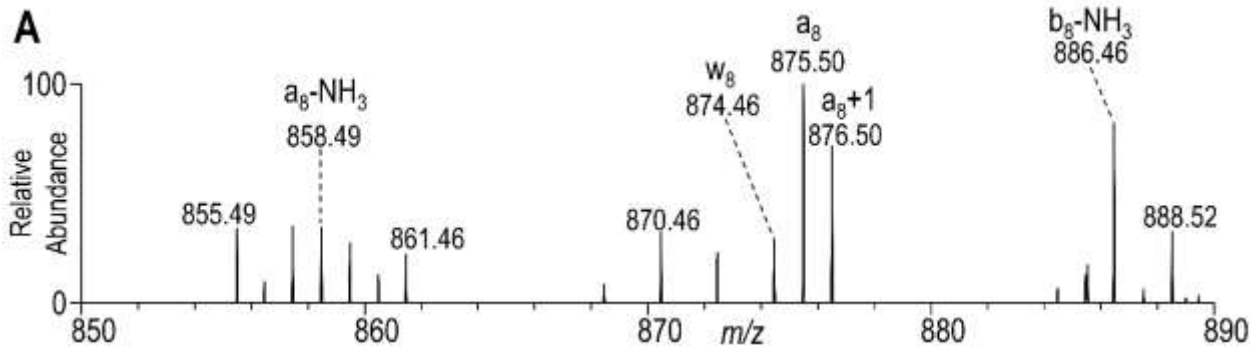
C

	Theoretical *m/z*	Observed *m/z*	Error (ppm)
AQY	363.1663	363.1659	1.10
x_3	371.1925	371.1924	0.27
a_4	372.2241	372.2237	1.07
a_4+1	373.2319	373.2316	0.87
b_4-H_2O	382.2085	382.2081	1.05
b_4-NH_3	383.1925	383.1921	1.04
b_5^{2+}	395.2292	395.2286	1.39
QYL-NH_3	388.1867	388.1863	1.03
YLI	390.2387	390.2382	1.28
b_7^{2+}	395.2292	395.2286	1.39
w_4	399.2238	399.2235	0.75

D

	Theoretical m/z	Observed m/z	Error (ppm)
w_4	399.2238	399.2235	0.75
b_4	400.2191	400.2188	0.75

Figure S5: Expansion of region around w_4 ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (2+ charge state) from the BB7.2 immunoprecipitation. Both (A) the lower *m/z* region as well as (B) a smaller excerpt around the b_4 ion are included because the high abundance of the b_4 ion makes it challenging to observe the other ions when presented using the same abundance scale. (C/D) Tables of the identified ions in A and B display the theoretical *m/z*, observed *m/z*, and ppm error. The mass of the w_4 ion that confirms the identification of leucine was calculated by subtracting 43.0548 Da from the mass of the z_4 ion.



B

	Theoretical m/z	Observed m/z	Error (ppm)
a_8-NH_3	858.4720	858.4686	3.96
w_8	874.4669	874.4644	2.86
a_8	875.4985	875.4964	2.40
a_8+1	876.5063	876.5034	3.31
b_8-NH_3	886.4669	886.4641	3.16

Figure S6: (A) Expansion of region around w_8 ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (1+ charge state) from the BB7.2 immunoprecipitation. (B) Table of the identified ions displays the theoretical *m/z*, observed *m/z*, and ppm error for each identified ion in the spectrum. All theoretical *m/z* values were calculated using Protein Prospector. The mass of the w_8 ion that confirms the identification of isoleucine was calculated by subtracting 43.0548 Da from the mass of the z_8 ion, and the mass of the a_8+1 ion was calculated by adding 1.0078 Da to the mass of a_3 ion.

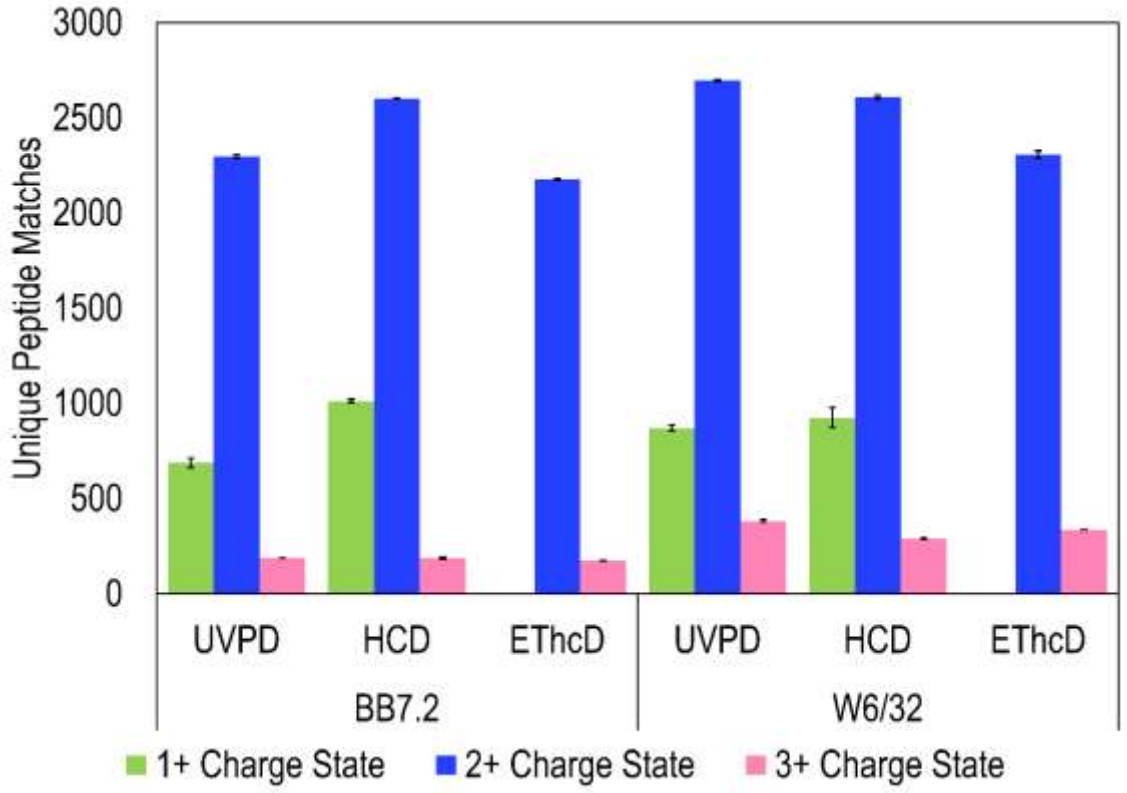


Figure S7: The number of peptide-spectrum matches (PSMs) for the BB7.2 and W6/32 immunoprecipitated samples categorized by charge state (1+, 2+, 3+) for UVPD, EThcD, and HCD mass spectra from Byonic output.

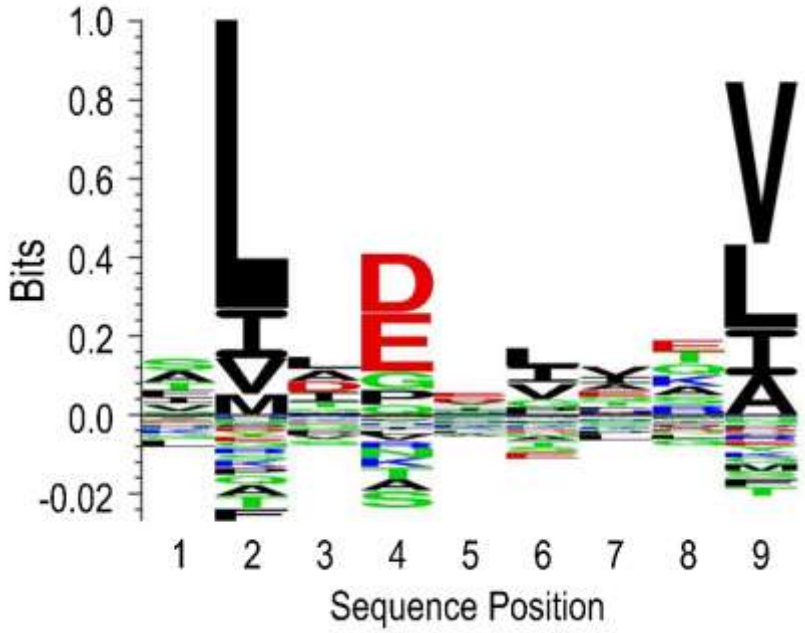


Figure S8: Sequence logo generated with Gibbs Cluster 2.0 from the 250 peptides used to produce Table 2.

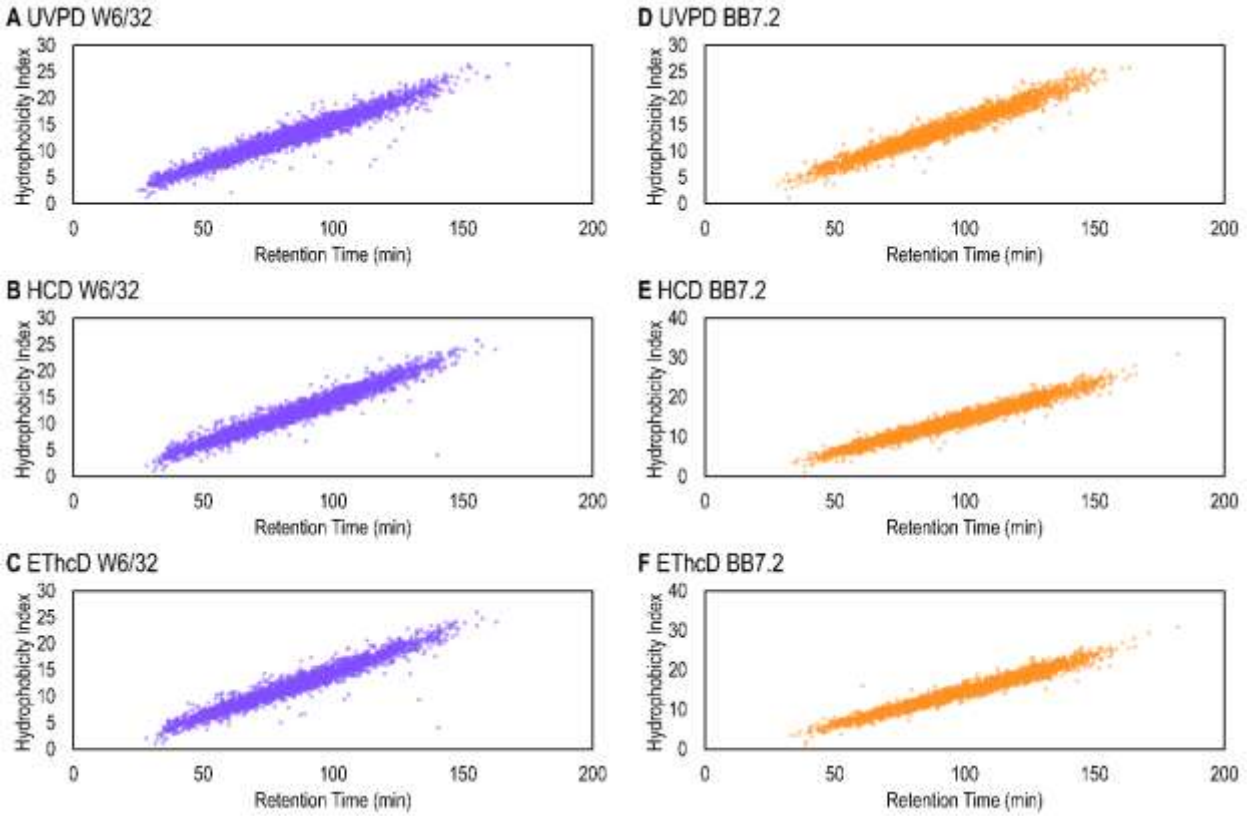


Figure S9: Plots of the retention time vs. the hydrophobicity index for the peptides in the W6/32 Immunoprecipitation, calculated using SSR calc for (A) UVPD, (B) HCD, and (C) EThcD spectra and in the BB7.2 immunoprecipitation for (D) UVPD, (E) HCD, and (F) EThcD spectra.

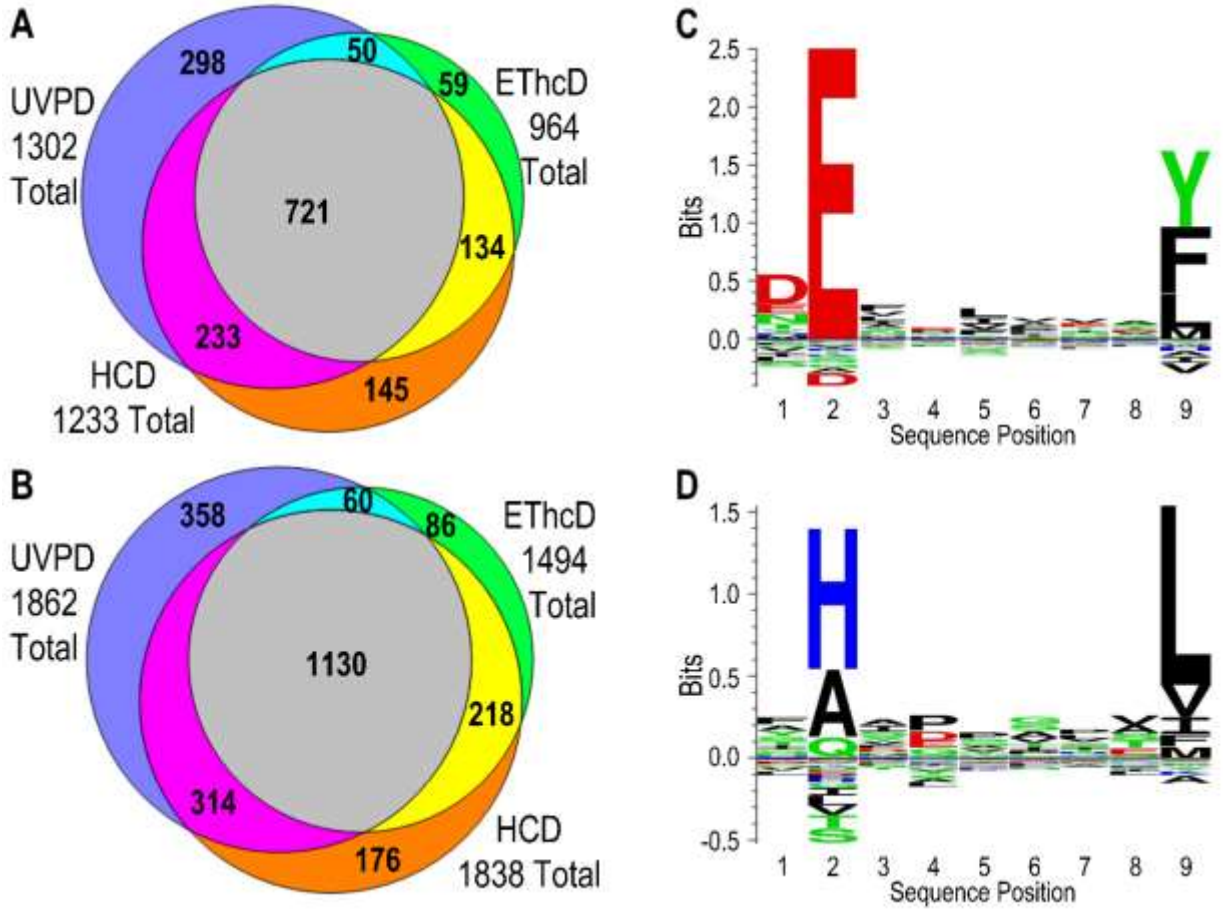
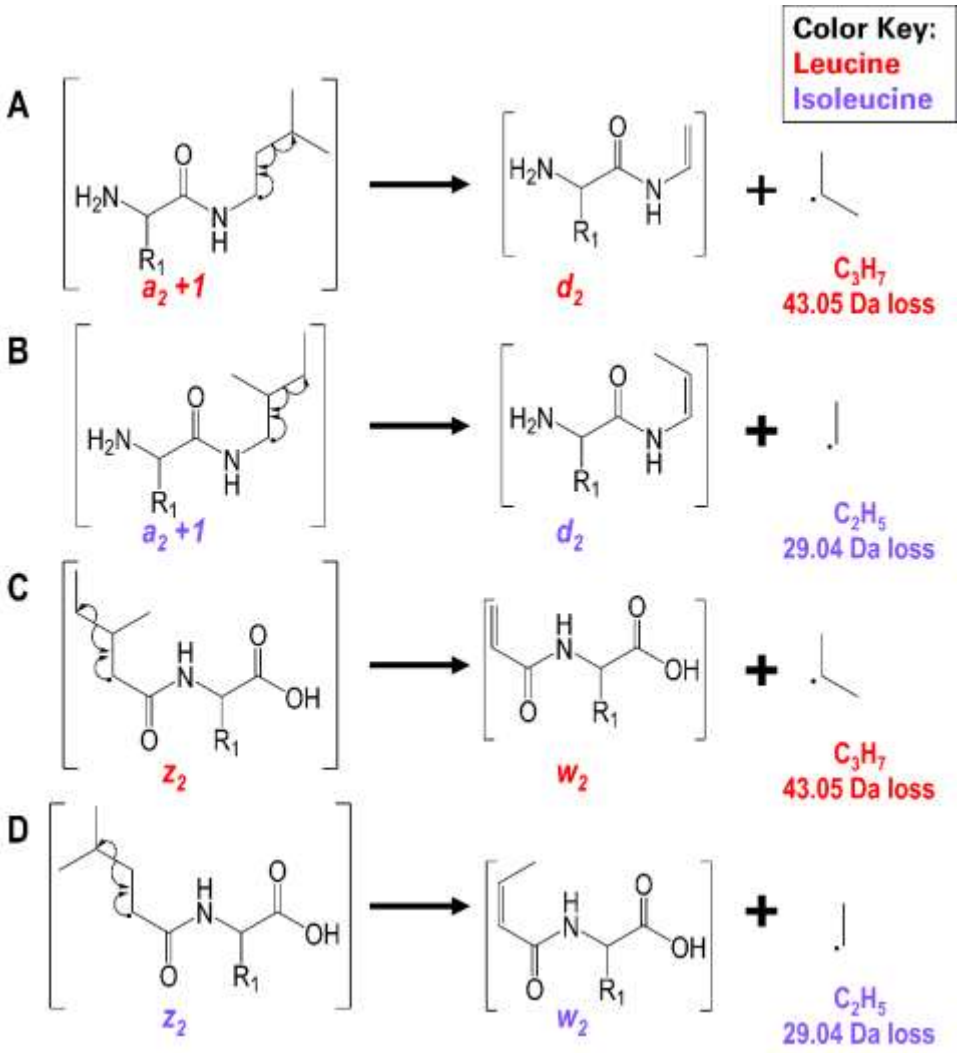


Figure S10: Venn diagrams displaying the total number of peptides identified utilizing UVPD, EThcD, and HCD for W6/32 immunoprecipitated sample. The peptides have been separated into either (A) HLA-B*15:10 or (B) HLA*18:01 based on cluster alignment in Gibbs Cluster 2.0. The sequence logos generated with Gibbs Cluster 2.0 are also included for both motifs (C) HLA-B*15:10 and (D) HLA*18:01 based on the combination of unique peptide matches across UVPD, EThcD, and HCD.



Scheme S1: Illustration of diagnostic neutral losses resulting in *d* and *w* ions that allow differentiation of leucine and isoleucine. The schema includes (A) a *d* ion resulting from the loss of C_3H_7 from a leucine-containing *a+1* ion, (B) a *d* ion resulting from the loss of C_2H_5 from an isoleucine-containing *a+1* ion, (C) a *w* ion resulting from the loss of C_3H_7 from a leucine-containing *z* ion, and (D) a *w* ion resulting from the loss of C_2H_5 from a leucine-containing *z* ion. The pathways are adapted from Xiao *et al*.[2]

Table S1: List of peptides included in the 157 synthetic peptide mixture (see Excel file)

	Number of isomeric amino acids	Number of isomeric amino acids confirmed	Percent confirmed	Number of *w* ions identified	Number of *d* ions identified
Leucine	736	130	18%	84	51
Isoleucine	322	38	12%	26	15

Table S2: Confirmation of I/L identities from BB7.2 immunoprecipitation samples achieved using 193 nm UVPD. All *d* and *w* ions were identified by Byonic and were only included if they appeared in duplicate LC runs.

	b/y	a/x and b/y	b/y and c/z	a/x, b/y, and c/z
UVPD	783	752	751	738
HCD	667	105	53	75

Table S3: The number of peptides identified from UVPD and HCD for W6/32 immunoprecipitation sample based on Comet searches including either just b/y, both a/x and b/y, both b/y and c/z, or all a/x, b/y, and c/z fragment ions. All peptides must be below a 1% q-value and appear in both replicates to be included.

	c/z	b/y and c/z	a/x and c/z	a/x, b/y, and c/z
EThcD	256	614	68	553

Table S4: The number of peptides identified from EThcD for W6/32 immunoprecipitation sample based on Comet searches including either just c/z, both b/y and c/z, both a/x and c/z, or all a/x, b/y, and c/z fragment ions. All peptides must be below a 1% q-value and appear in both replicates to be included.

Table S5: List of peptides identified in the W6/32 immunoprecipitation. Every peptide has one entry for every MS/MS method for which it was successfully identified. Also included is the retention time associated with the scan that produced the highest score in Byonic for the first replicate, along with the Hydrophobicity Index as calculated with SSR calc. These two pieces of information were used to generate Figure S9. The highest reported Byonic Scores are included for each replicate, along with any post-translational modifications associated with the scan resulting in each score. (see Excel file)

Table S6: List of peptides identified in the BB7.2 immunoprecipitation. Every peptide has one entry for every MS/MS method for which it was successfully identified. Also included is the retention time associated with the scan that produced the highest score in Byonic for the first replicate, along with the Hydrophobicity Index as calculated with SSR calc. These two pieces of information were used to generate Figure S9. The highest reported Byonic Scores are included for each replicate, along with any post-translational modifications associated with the scan resulting in each score. For UVPD and EThcD any *d/w* ions identified are also listed. These *d/w* ions were all identified by Byonic in both replicates and are only included if they represent a neutral loss that supports the expected result consistent with the Uniprot database. (see Excel file)

References:

(1) Potts, G. K.; Ready, D. B.; George Thompson, A.; Williams, J.; Patterson, M. An Optimized Method For Achieving Maximal HLA-A*02:01 Immunopeptide Identifications From Single Shot Analysis On An Orbitrap Lumos Mass Spectrometer. In *ASMS 2018 San Diego*; ASMS: San Diego, CA, 2018; Vol. 66.

(2) Xiao, Y.; Vecchi, M. M.; Wen, D. Distinguishing between Leucine and Isoleucine by Integrated LC–MS Analysis Using an Orbitrap Fusion Mass Spectrometer. *Anal. Chem.* **2016**, *88*, 10757–10766.