SUPPORTING INFORMATION

We also examined the hydrogen transfer process in ETD which generates c-1 and z+1 ions. A deisotoping procedure was used to remove the portion of the z+1 ion intensity that comes from the first isotope peak of a regular z ion. Both c-1 and z+1 ions were mined for clustering. For the Lys-C (Figure 6) and Glu-C (Figure 7) datasets, strong c-1 ions were observed for cleavage N-terminal to Lys and Glu, respectively, while no strong z+1 ions were observed. We examined the sequence features (charge, length, residue positions, etc.) for this separation and did not see any dominating factor. This is in conflict with previous reports that extensive hydrogen transfer products were observed for doubly charged peptides, but much less for triply or higher charged peptides.[12] The tryptic dataset does not contain doubly charged peptides so Figure 8 only reflects pairwise fragmentation patterns of c-1 and z+1 ions for peptides with three and more charges. Clustering produced two clusters, one with no c-1, z+1 ions and the other with very strong c-1 ions at the N-terminus of Lys and relatively strong z+1 ions at the N terminus of Arg. The clusters for c-1 and z+1 ions from the three datasets share the same trends as c and z ions, which may imply that hydrogen transfer products are coming from the same dissociation pathway as c, z ions. Another possibility is that polar C-terminal residues (Glu, Lys) may stabilize the (c + z) complexes and facilitate exothermic H-atom migration.

List of CART features, where "count" means a simple count of the number of that residue in a sequence, POS means the position of the AA residue in the sequence, indicated as a fraction, DistN means the distance between N terminus and a certain residue, while DistC

means the distance between C-terminus and a certain residue, m/z means mass-to-charge ratio, and H Mobile is defined as (charge – count.R- 0.5*count.K - 0.5* count.H): charge, count.Basic, count.Acidic,count.A, count.C, count.D, count.E, count.F, count.G, count.H, count.I, count.K, count.L, count.M, count.N, count.P, count.Q, count.R, count.S, count.T, count.V, count.W, count.Y, POS.A, POS.C, POS.D, POS.E, POS.F, POS.G, POS.H, POS.I, POS.K, POS.L, POS.M, POS.N, POS.P, POS.Q, POS.R, POS.S, POS.T, POS.V, POS.W, POS.Y, DistN.A, DistN.C, DistN.D, DistN.E, DistN.F, DistN.G, DistN.H, DistN.I, DistN.K, DistN.L, DistN.M, DistN.N, DistN.P, DistN.Q, DistN.R, DistN.S, DistN.T, DistN.V, DistN.W, DistN.Y, DistC.A, DistC.C, DistC.D, DistC.E, DistC.F, DistC.G, DistC.H, DistC.I, DistC.K, DistC.L, DistC.M, DistC.N, DistC.P, DistC.Q, DistC.R, DistC.S, DistC.T, DistC.V, DistC.W, DistC.Y, Length, m/z, H.Mobile

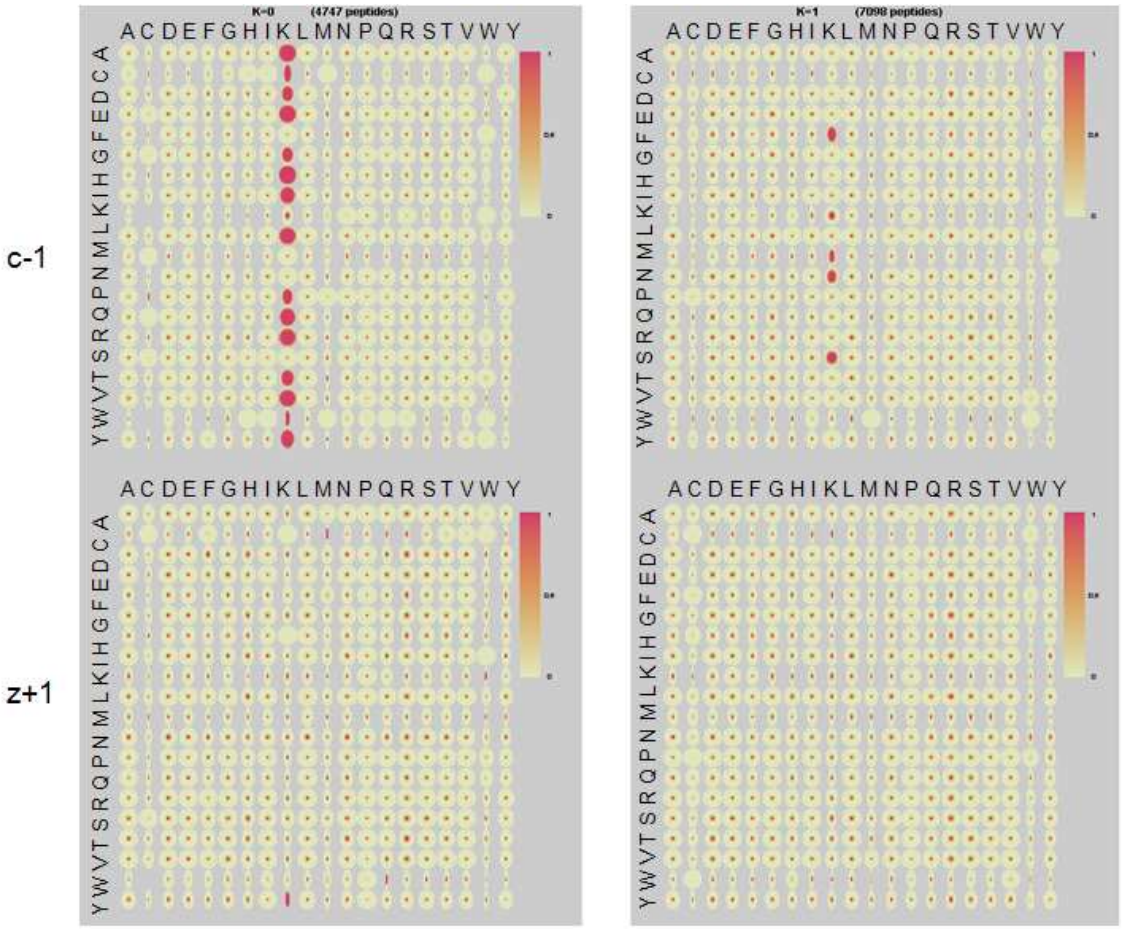

Figure 6. Quantile maps for the two clusters obtained by K-means clustering of 11954 spectra of Lys-C unique peptides, for c-1 (top) and z+1 ions (bottom).

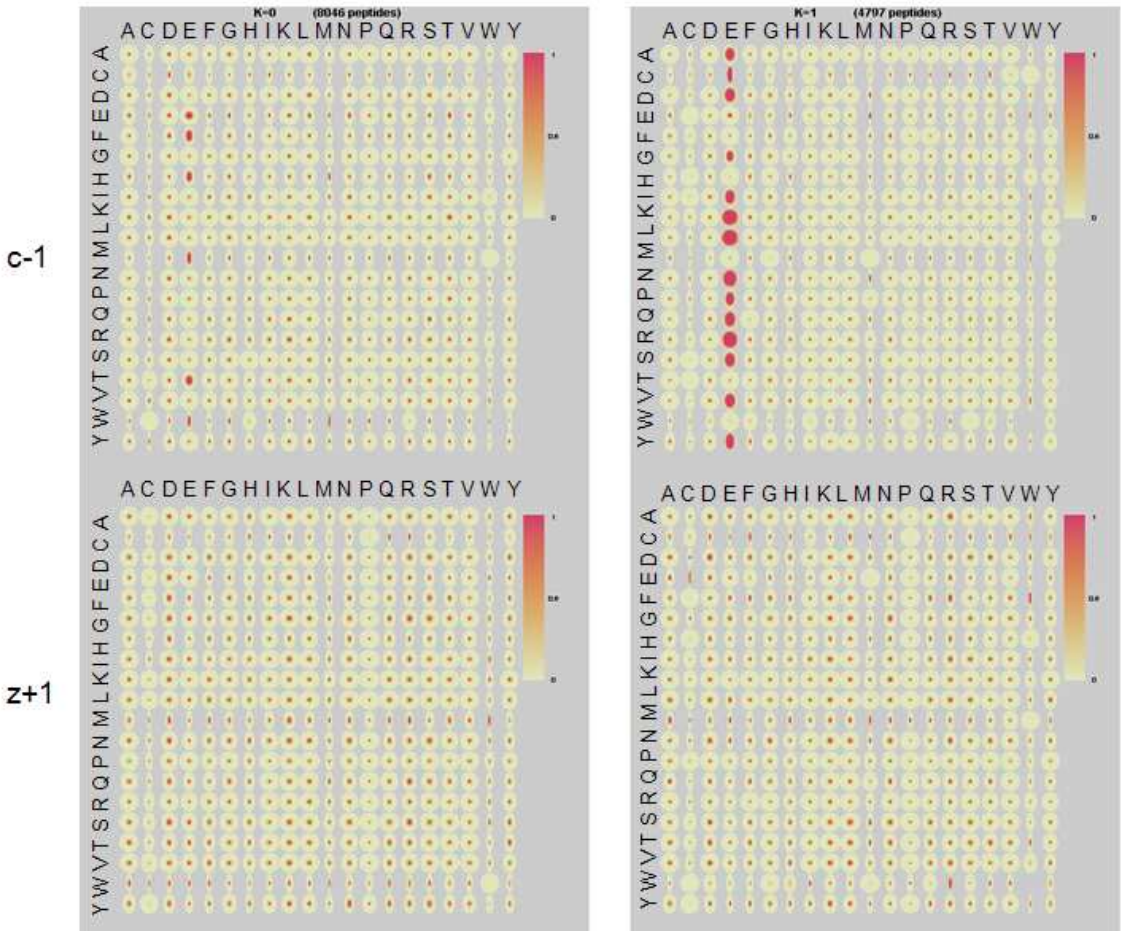

Figure 7. Quantile maps for the two clusters obtained by K-means clustering of 12042 spectra of Glu-C unique peptides, for c-1 (top) and z+1 ions (bottom).

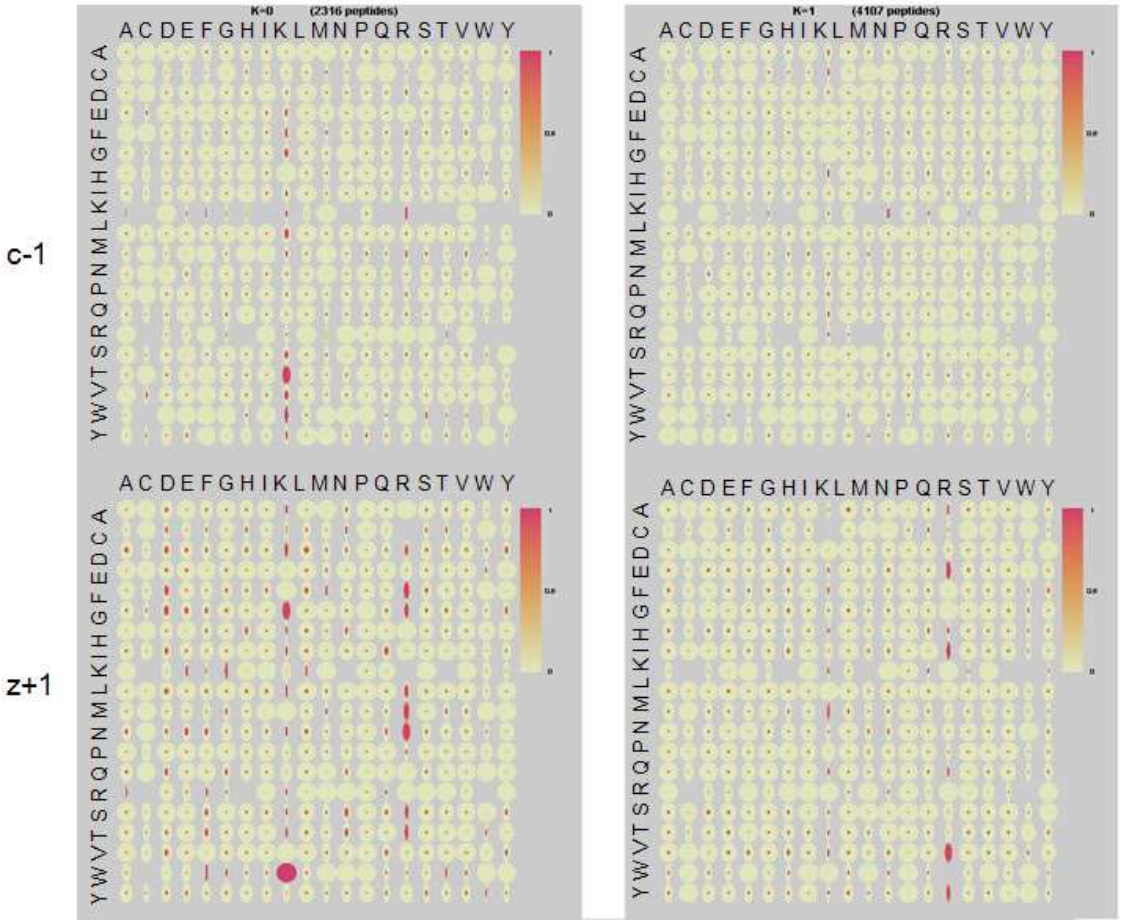

Figure 8. Quantile maps for the two clusters obtained by K-means clustering of 6423 high resolution spectra of tryptic unique peptides, for c-1 (top) and z+1 ions (bottom).