Supporting information

Site-Selective Modification of Proteins Using Cucurbit[7]uril as Supramolecular Protection for N-terminal Aromatic Amino Acids

Anton A. A. Smith,[a,b] Caitlin L. Maikawa,[a] Gillie A. Roth[a] and Eric A. Appel [c]

Materials

All reagents were acquired through Millipore-Sigma.

Insulin modification and digestion

An insulin stock solution of 4 mg/mL was prepared by dispersing 8 mg of insulin in 0.5 mL milliQ water and acidifying with 1 μL aq HCl (conc) until the dispersed powder dissolved. 0.5 mL 2x PBS pH 7.3 buffer was added, resulting in a temporarily cloudy precipitate that quickly dissipates.

Cucurbit[7]uril (10 μL, 6.8 mM or 3.7 mM, solution in PBS) was added to insulin (20 μL, 0.69 mM, 4 mg/mL). This resulted in an opaque suspension believed to be Insulin complexed with CB[7]. This effect was not due to a drop in pH beyond buffer capacity. m-dPEG$_8$-NHS ester was added from a freshly prepared stock solution (8 μL, 3.9 mM or 84.4 mM). This solution was incubated for 37 °C for 2 hours. Glu-C digestion was performed by adding Glu-c (8 μL, 1 mg/mL) and incubated at 37 °C for a minimum of 4 hours. DTT reduction of non-digested insulin was performed by adding 10 μL 5 mM DTT, instead of Glu-c. Solutions for MS where prepared by 100x dilution into 1:1 ACN:water.

LC-MS Methods

Samples were analyzed by LC-ESI/MS on a Waters Acquity UPLC and SQD2 single quadrupole mass spectrometer. A Zorbax 300SB-C8 2.1x50mm 3.5u HPLC column was used with the following gradient:

Time (mins)	**Flow rate (mL/min)**	**Solvent A (0.1% formic acid in water)**	**Solvent B (0.1% formic acid in acetonitrile)**
0	0.3	98	2
2	0.3	98	2
6	0.3	5	95
8	0.3	5	95
8.5	0.3	98	2
9.5	0.3	5	95
10.5	0.3	5	95
11	0.3	98	2
15	0.3	98	2

Data was collected in positive ion, full scan MS mode with a mass range of 50-2000 Da. High-resolution masses were determined on a Waters Acquity UPLC and Thermo Exactive Orbitrap Mass Spectrometer. An Agilent Poroshell C18 2.1x50mm 2.7μm HPLC column, with an identical gradient as above, was used for separation. Data was collected in positive ion, full scan MS mode with a mass range of 100-2000 Da, Maximum ion injection time 250 ms, Resolution=50K, AGC=1x10^6.

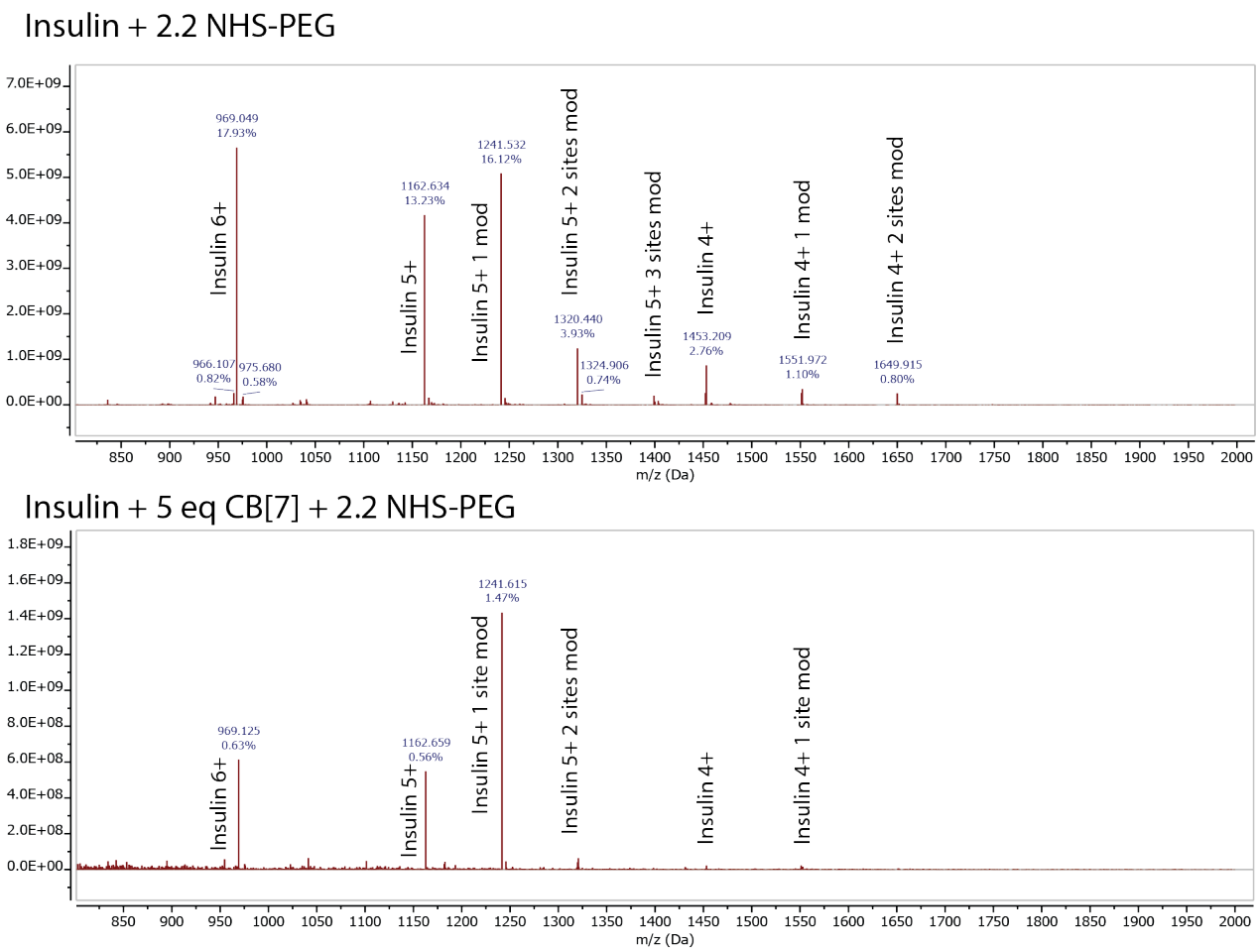

Figure S1: Low resolution mass spectrometry of whole insulin modified with 2.2 eq NHS-PEG ester in the absence (top) or presence (bottom) of CB[7].

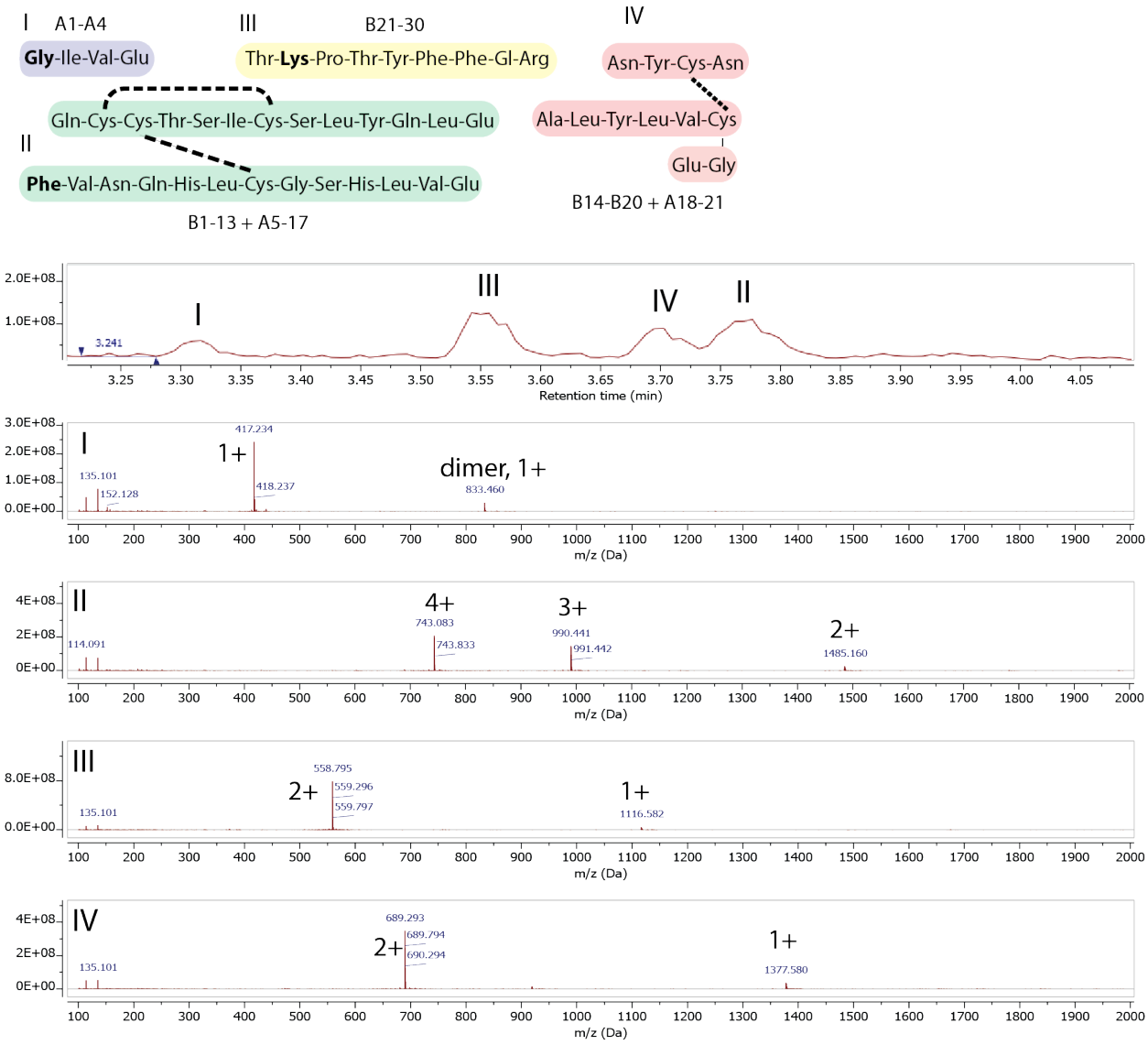

Figure S2: Glu-C digested insulin with identified fragments, high resolution. Top bar shows ion count intensity elugram, with masses corresponding to each of the fragments shown in the lower 4 panels.

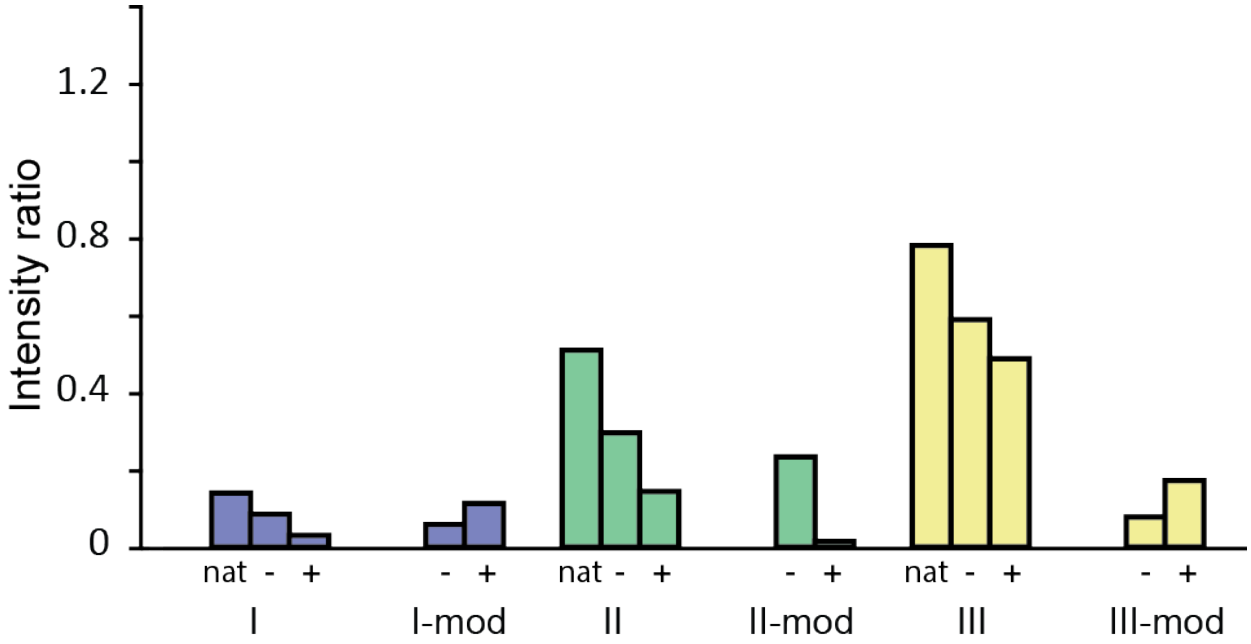

Figure S3: Intensity ratios of fragments, relative to fragment IV. Either native unmodified insulin, insulin modified with 2.2 equivalents m-dPEG$_8$-NHS ester, with or without 5 eq. of CB[7] added. In the presence of CB[7], I and III is modified to a greater extent. II appears to be almost completely non-modified, but shows an overall decreased intensity in the presence of CB[7], likely as a consequence of complexation with CB[7]. Each fragment intensity is from all isotopes and m/z from Figure S6 and Figure S7

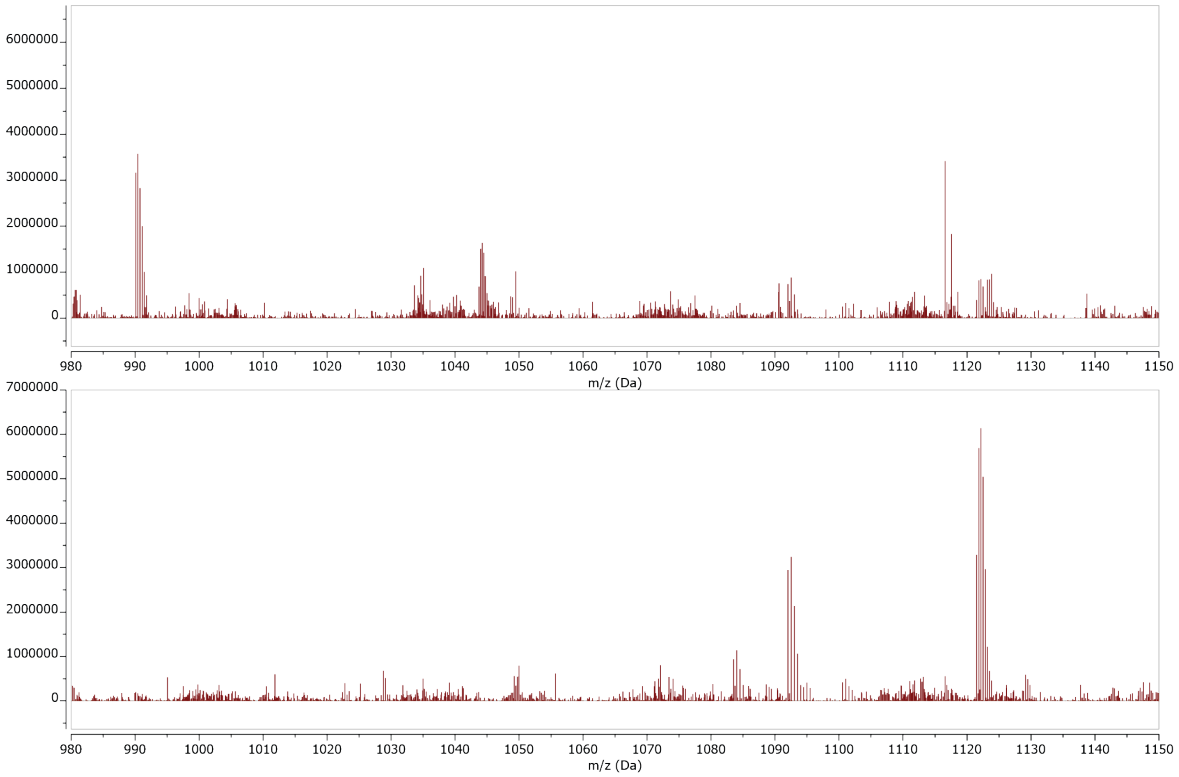

Figure S4: Insulin incubated with 50 eq. of m-dPEG-NHS ester, in the presence of 2.7 equivalents of CB[7] (top) or absence of CB[7] (bottom). In the absence of CB[7], the $+3H^+$ ion (m/z 990) from fragment II is greatly reduced, while CB[7] reduces the $+3H^+$ ion of the modified fragment II (m/z 1122) .

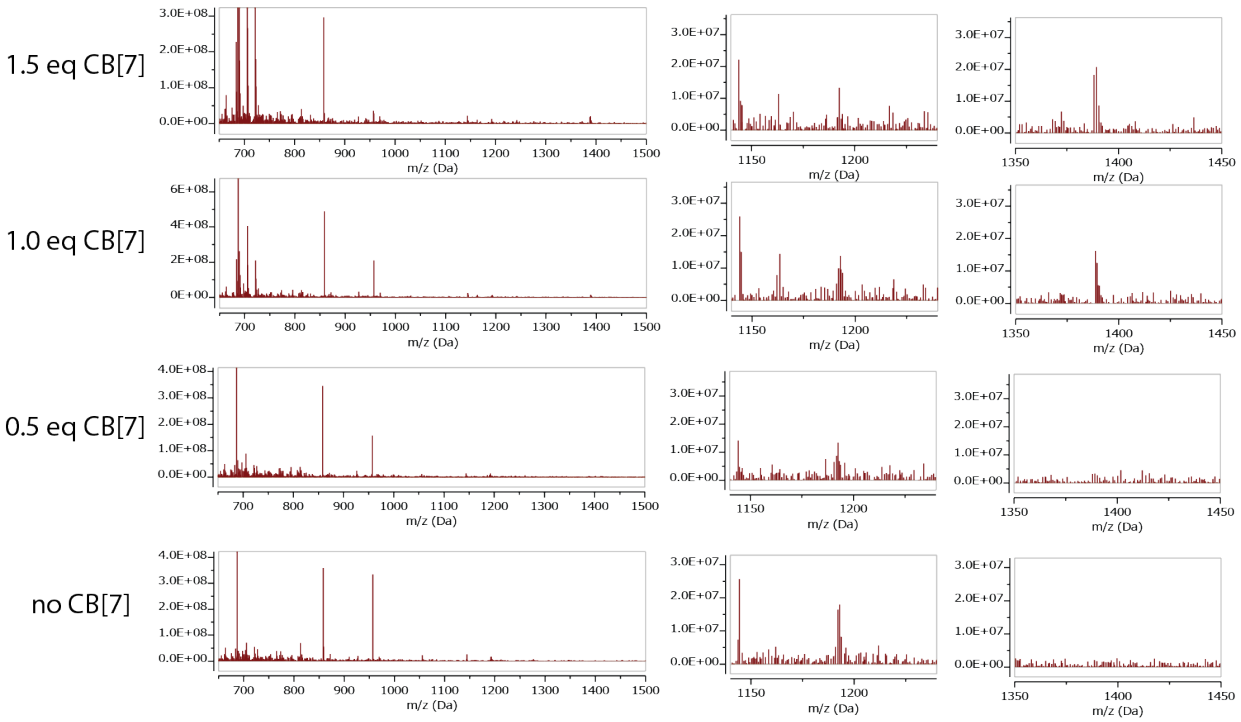

Figure S5: Various amounts of CB[7] with insulin and 1.5 eq m-dPEG$_8$-NHS, followed by a reduction with DTT. m/z 1192 is the native A chain +2H$^+$ (monoisotopic mass = 2381.989u), with the modified fragment found at m/z 1387. m/z 858.42 is the +4H$^+$ ion of the B chain (monoisotopic mass = 3427.67u), with the modified fragment found at m/z 955.9. Lower intensities are seen for the A chain, due to a lack of multiple positively charged amino acids. An increase in A chain functionalization is seen with increasing CB[7], with a corresponding decrease in B chain functionalization.

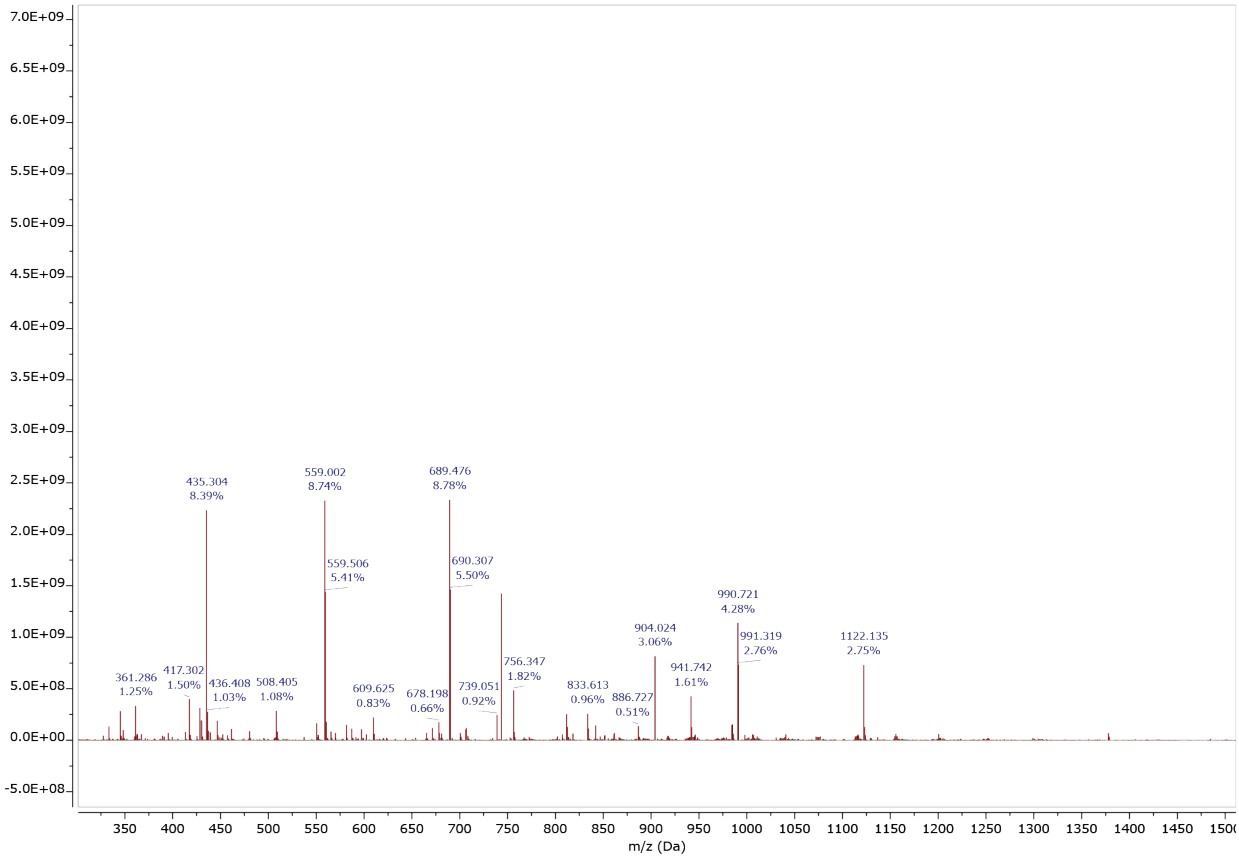

Figure S6: Insulin + 2.2 eq m-dPEG$_8$-NHS, digested. m/z 417.3 Fragment I +H$^+$, m/z 435.3 m-dPEG$_8$-COOH, m/z 559.0 Fragment III +2H$^+$, m/z 689.5 Fragment IV +2H$^+$, m/z 743.1 Fragment II modified +4H$^+$, m/z 756.4 Fragment III +2H$^+$, m/z 811.6 Fragment I modified +H$^+$, m/z 833.6 Fragment I dimer +H$^+$, m/z 990.7 Fragment II +3H$^+$, m/z 1122.1 Fragment II +3H$^+$.

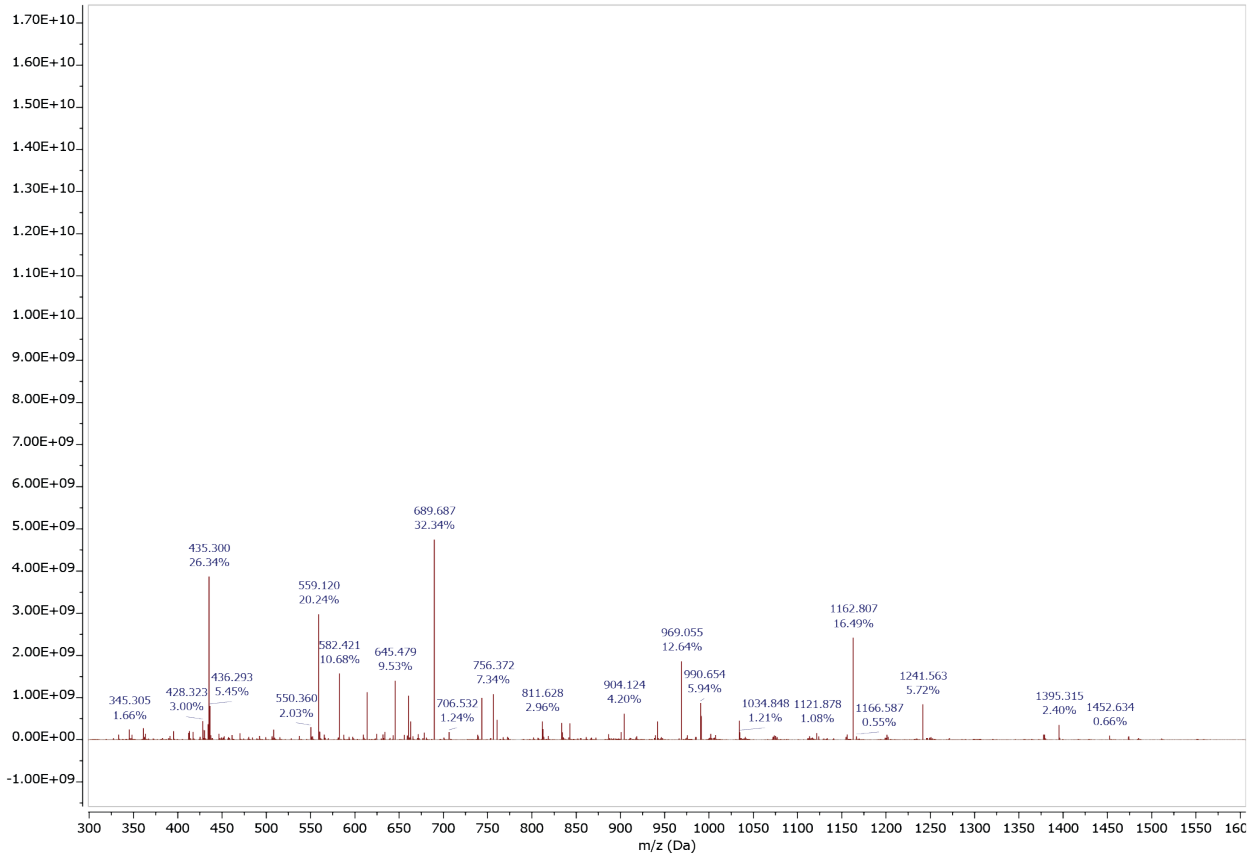

Figure S7: Insulin + 5 eq CB[7] + 2.2 eq m-dPEG$_8$-NHS. Insulin + 2.2 eq m-dPEG$_8$-NHS, digested.m/z 435.3 m-dPEG$_8$-COOH, m/z 559.0 Fragment III +2H$^+$, m/z 689.5 Fragment IV +2H$^+$, m/z 743.1 Fragment II modified +4H$^+$, m/z 756.4 Fragment III +2H$^+$, m/z 811.6 Fragment I modified +H$^+$, m/z 969.1 Insulin +5H$^+$, m/z 1162.8 Insulin +6H$^+$, m/z 990.7 Fragment II +3H$^+$, m/z 1122.1 Fragment II +3H$^+$, m/z 1241.6 insulin 1 mod +5 H$^+$, m/z 1395.3 insulin+CB[7] +5 H$^+$.

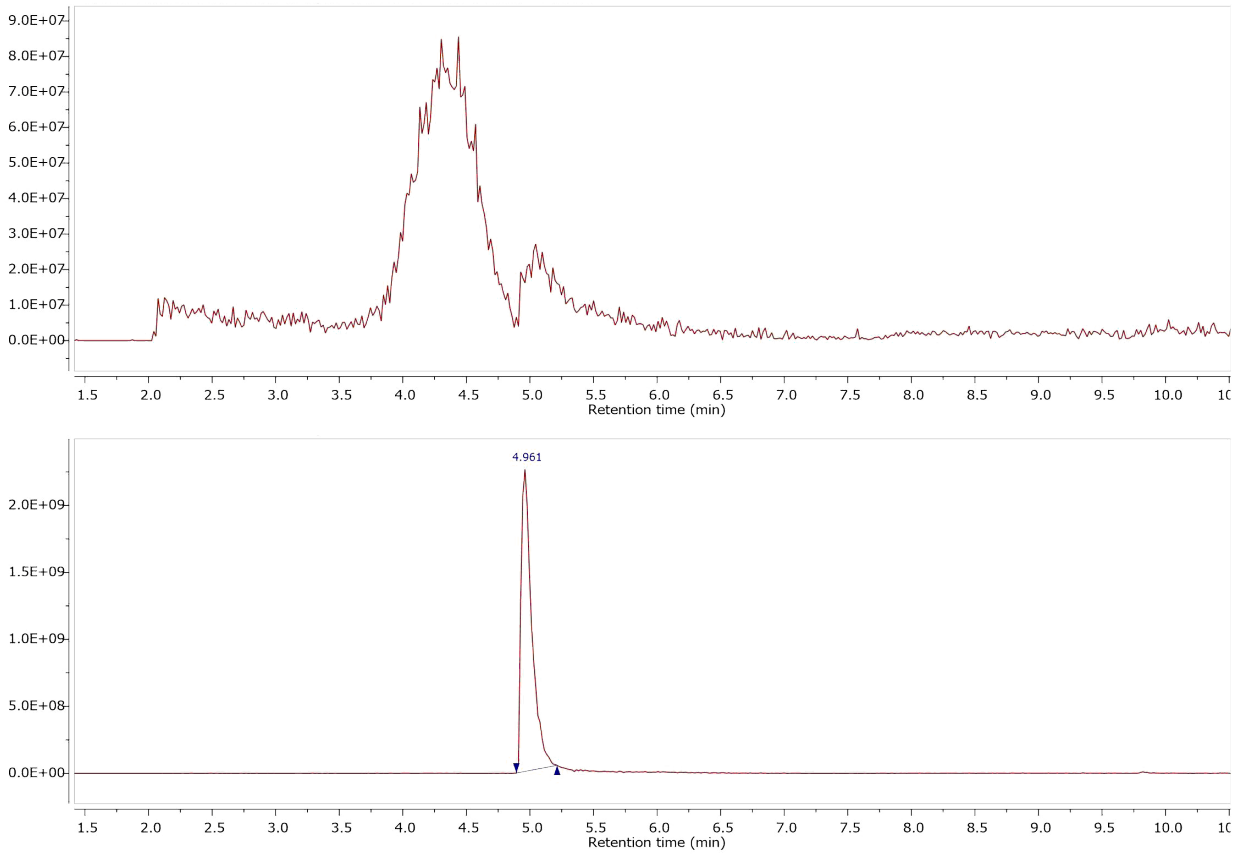

Figure S8: (top) TIC elugram for m/z associated with CB[7]. (bottom) m/z 582 TIC elugram for m/z associated with intact insulin, m/z 958. The different elution times show that HPLC can sufficiently separate the non-covalent complex between insulin and CB[7]