Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Information

Table of Contents

1) Chemical Syntheses and Analysis

1.1 General Information

All commercially available reagents were used without further purification. 3,5-Dihydroxybenzyl alcohol, carbon tetrabromide, triphenylphosphine, sodium azide, and *N*-hydroxysuccinimide (NHS) were purchased from Acros Organics (Geel, Belgium). Ethyl 7-bromoheptanoate and human serum albumin (HSA) were obtained from Sigma-Aldrich (Tokyo, Japan). 1-Ethyl-3-(3-(dimethylamino)propyl)carbodiimide hydrochloride (EDC) was provided by TCI chemicals (Tokyo, Japan). All *N*-glycans were supplied from Glytech (Kyoto, Japan). For the fluorophores used in this study, HiLyte^{TM} Fluor750 acid SE® was purchased from AnaSpec (Fremont, USA), and 5-(and-6)-Carboxytetramethylrhodamine NHS ester was purchased from Molecular Probes (Oregon, USA).

Ultrapure water used for all synthetic experiments described in this paper was obtained from a Milli-Q Advantage® A10 Water Purification System sold by Merck Millipore (Burlington, USA). In addition, Amicon 10K® centrifugal filters and Durapore PVDF 0.45 μm® filters were also purchased from by Merck Millipore (Burlington, USA). For chemical synthesis, high-resolution mass spectra (HRMS) were obtained on a Bruker micrOTOF-QIII spectrometer® by electron spray ionization (ESI-TOF-MS). For glycan-conjugated albumins, matrix assisted laser desorption ionization (MALDI-TOF) mass spectrometry analysis was obtained on a Bruker autoflex spectrometer® using 2,5-dihydroxybenzoic acid as a matrix.

Reverse phase HPLC analysis/purification was performed on a Shimadzu Prominence® system equipped with a Nacalai tesque column (5C18-AR-300, 4.6 x 250 mm). Two solvent systems, A: H_2O containing 0.1% TFA and B: MeCN containing 0.1% TFA, were applied.

1.2 Preparation of Glycan–Aldehyde probes

Glycan–aldehyde probes were prepared according to the previously reported procedure.[1] Detailed structures and simplified notations are given in Figure S1.

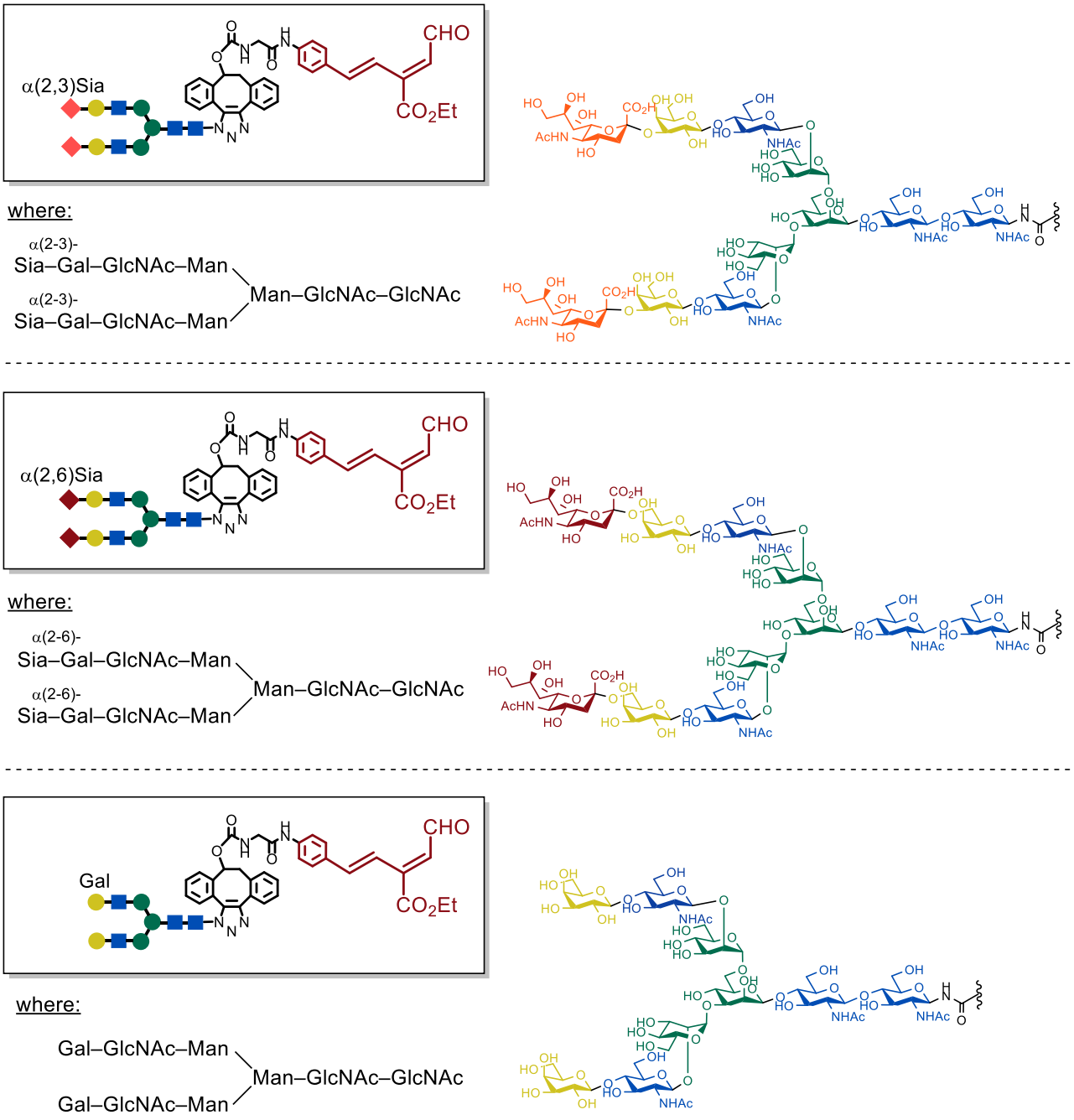
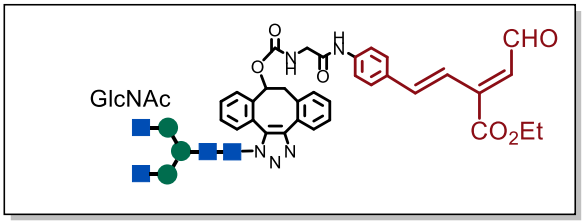
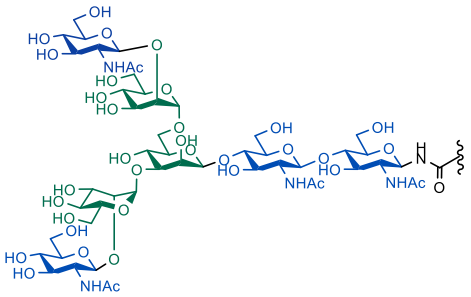


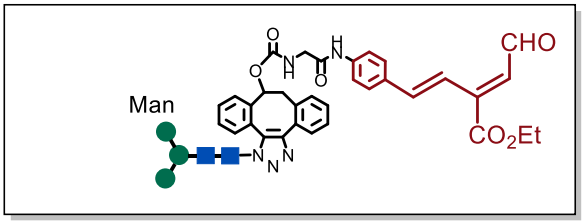
Figure S1. Whole and simplified chemical structures of the glycan–aldehyde probes used in this study



where:

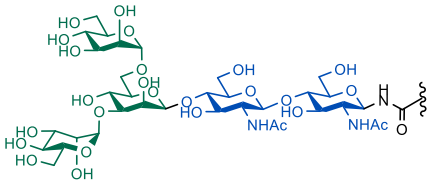
GlcNAc–Man
Man–GlcNAc–GlcNAc
GlcNAc–Man

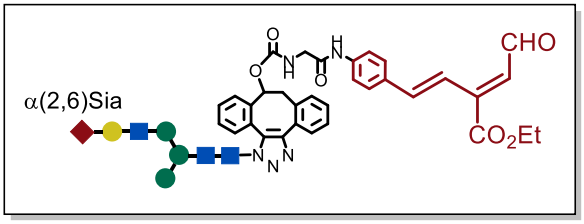




where:

Man
Man–GlcNAc–GlcNAc
Man





where:

α(2-6)-
Sia–Gal–GlcNAc–Man
α(1-6)-
Man–GlcNAc–GlcNAc
Man

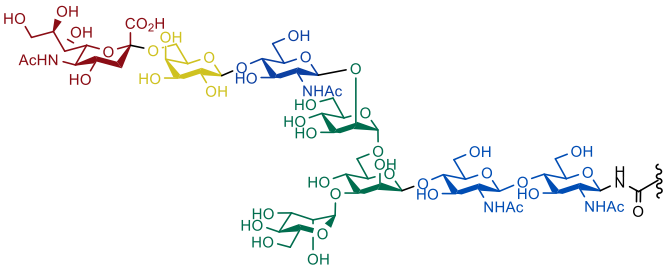
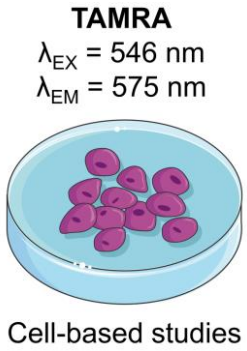


Figure S1 (cont). Whole and simplified chemical structures of the glycan–aldehyde probes used in this study

1.3 Glycoalbumins for Cell- and Animal-based Studies

1.3.1 Preparation of Fluorescently Labeled Albumin

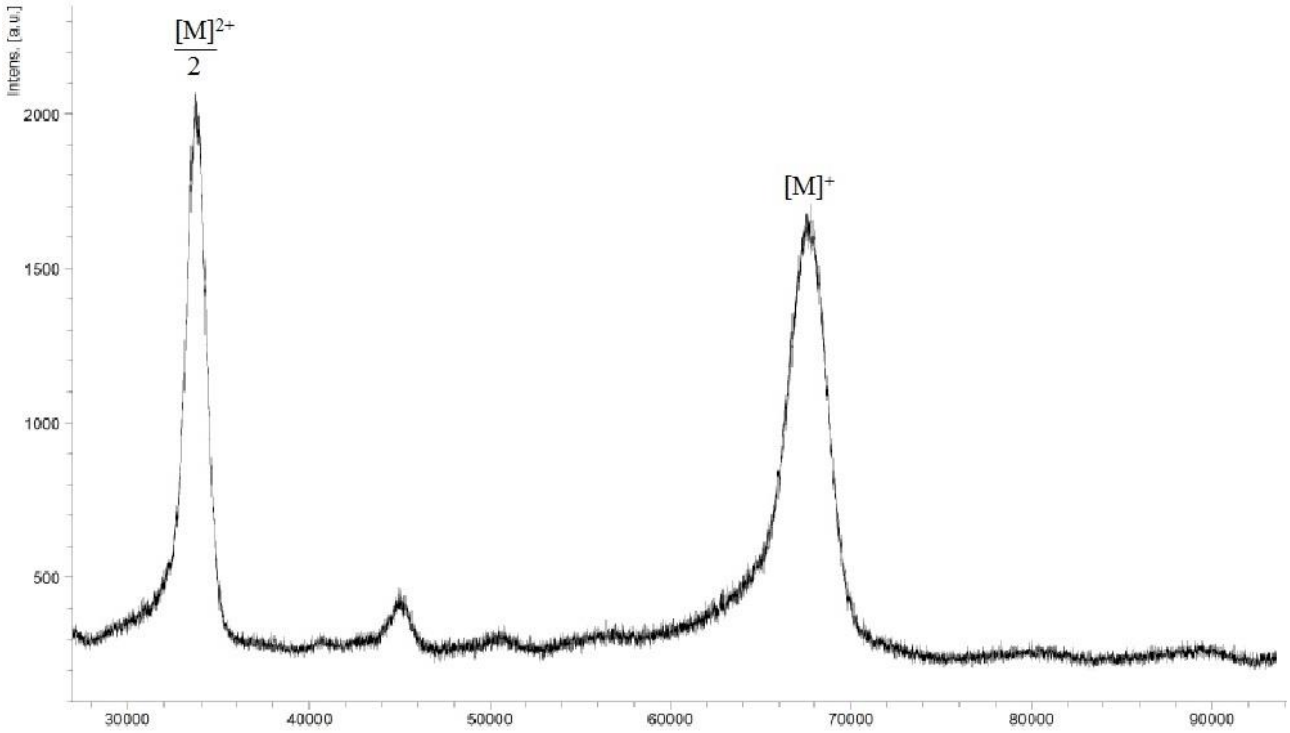


Cell-based studies

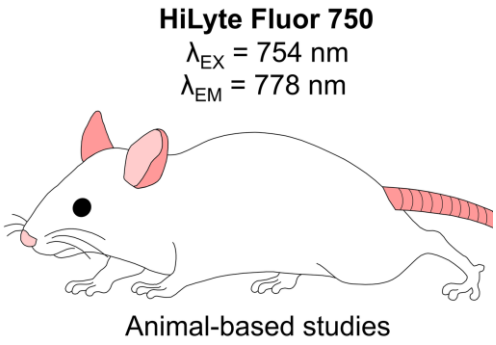
For cell-based studies, glycoalbumins **2a-f** were labeled with the fluorescent TAMRA dye. In this document, these are denoted as **TAMRA-2a**, **TAMRA-2b**, **TAMRA-2c**, **TAMRA-2c'**, **TAMRA-2d**, **TAMRA-2e**, and **TAMRA-2f**.

Preparation of **TAMRA-HSA**

To a solution of human serum albumin (HSA, purchased from SIGMA, 6.7 mg, 100 nmol) in phosphate saline buffer (pH = 7.4, 1.00 mL) was added TAMRA SE (0.13 mg, 250 nmol) in DMSO (13 μL) and the mixture was warmed to 37 °C. After 15 min, the solution was centrifuged through Amicon 10K® at 15,000 rpm for 10 min to filter off the small molecules. The residue was further washed with water and centrifuged for four times. The resulting solution was diluted by ultrapure water to 800 μL to afford the stock solution of **TAMRA-HSA**, which was used for subsequent *N*-glycan modifications. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-HSA** at 67.9 kDa, which contains an average number of 2 TAMRA fluorophores per albumin molecule.



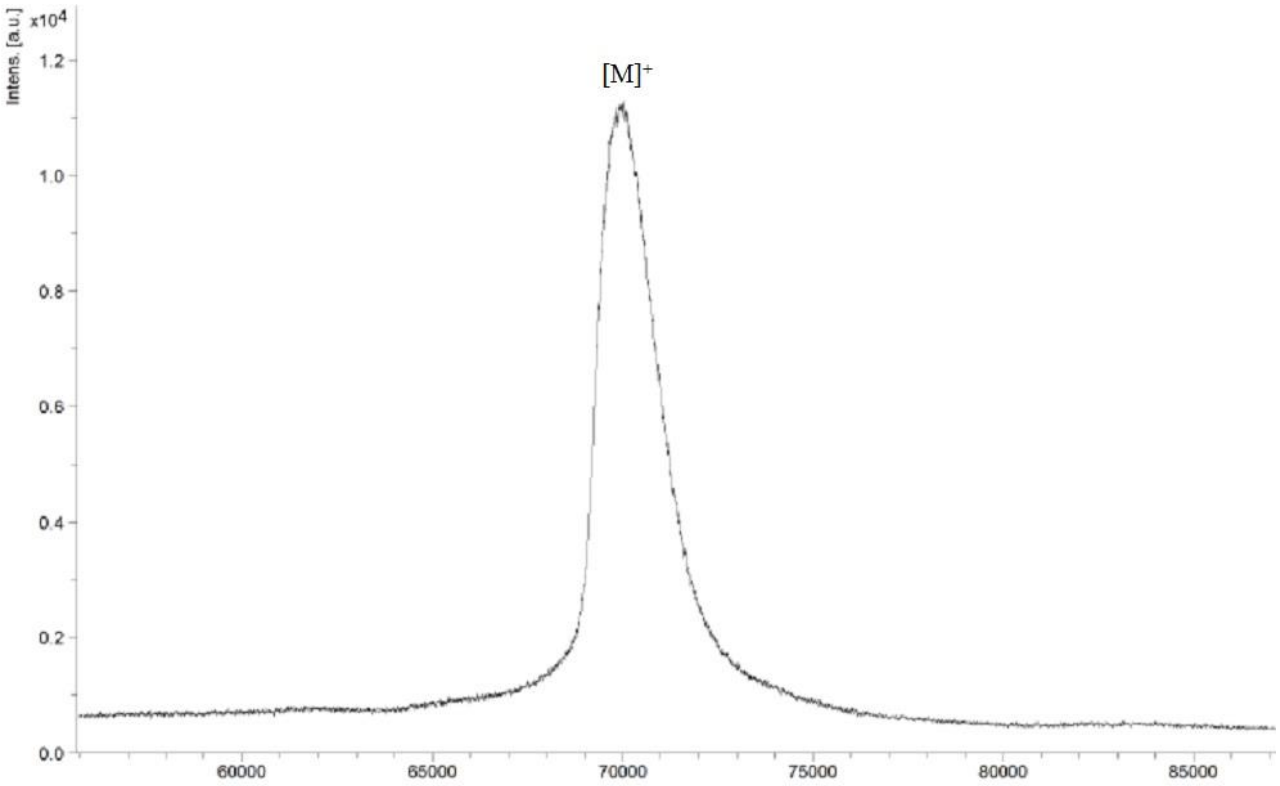
MALDI-TOF-MS of **TAMRA-HSA**



For animal-based studies, glycoalbumins **2a-f** were instead labeled with the near-infrared dye, HiLyte Fluor 750. In this document, these are denoted as **HiLyte-2b**, **HiLyte-2c**, and **HiLyte-2e**.

Preparation of **HiLyte-HSA**

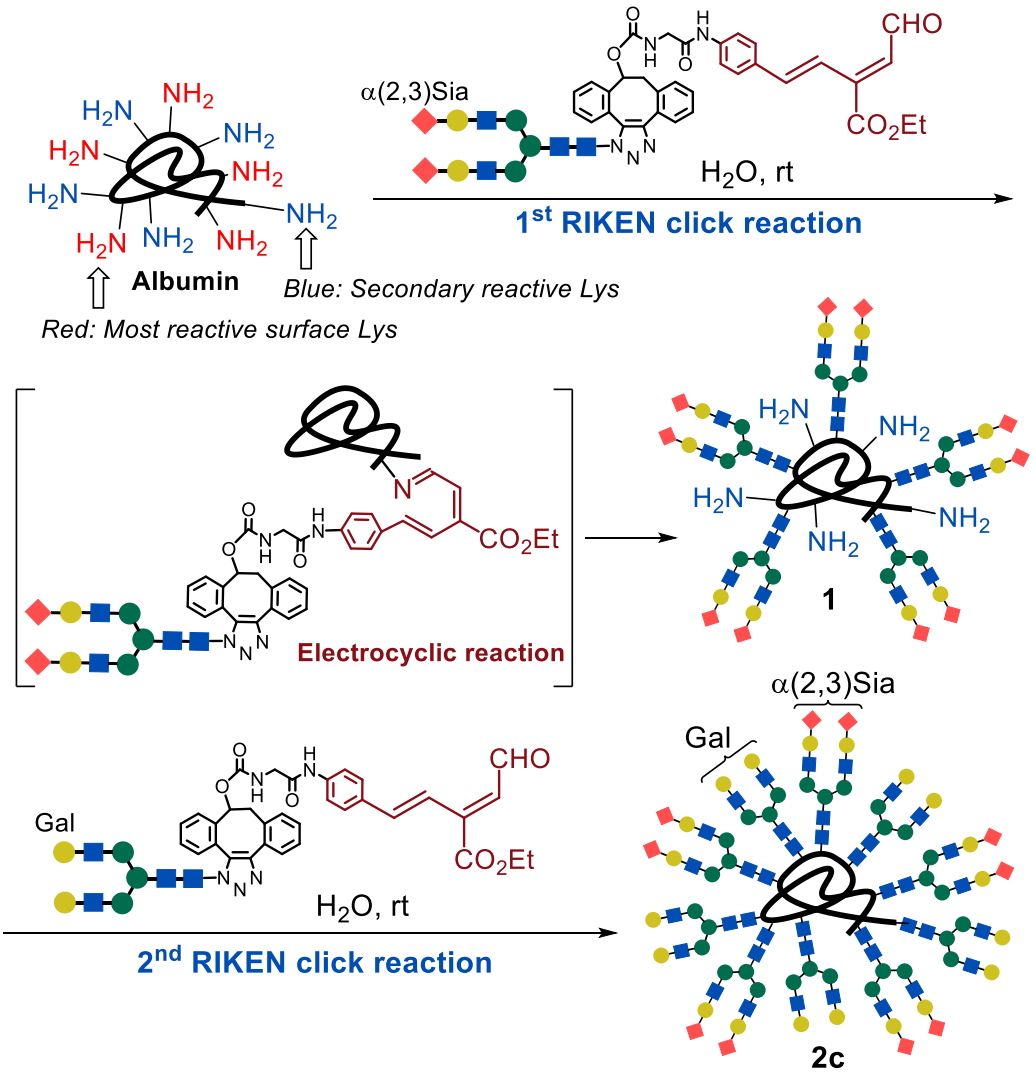
To a solution of human serum albumin (HSA, purchased from SIGMA, 4.8 mg, 72 nmol) in phosphate saline buffer (pH = 7.4, 720 μL) was added HiLyte Fluor 750 acid SE® (0.38 mg, 0.29 μmol, tetraethylammonium salt) in DMSO (10 μL) and the mixture was warmed to 37 °C. After 10 min, the solution was centrifuged through Amicon 10K® at 15,000 rpm for 10 min to filter off the small molecules. The residue was further washed with phosphate buffer and centrifuged for three times. The resulting solution was diluted by ultrapure water to 1.14 mL to afford the stock solution of **HiLyte-HSA**, which was used for subsequent *N*-glycan modifications. MALDI-TOF-MS (positive mode) detected the molecular weight of **HiLyte-HSA** at 70.3 kDa, which contains an average number of 2.9 HiLyte Fluor 750fluorophores per albumin molecule.



MALDI-TOF-MS of **HiLyte-HSA**

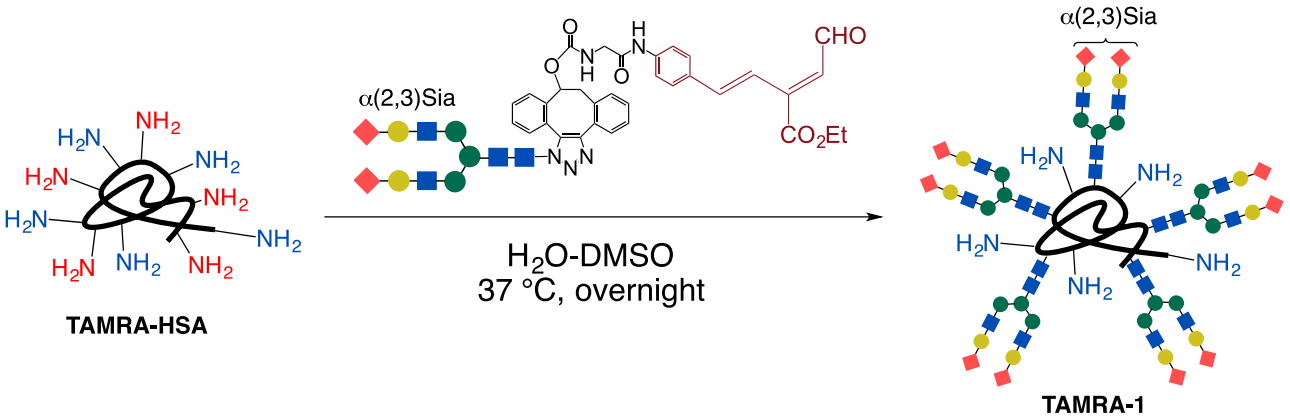
1.3.2 Preparation of Glycoalbumins **2a-f**

In the general synthetic scheme of glycoalbumins **2a-f** (Scheme S1), preparation was carried out in a sequential two-step manner. To begin, fluorophore-labeled HSA (either **TAMRA-HSA** or **HiLyte-HSA**) was treated with various glyco-aldehyde probes to create the intermediary glycoalbumin, which is generally conjugated to about 5 glycan moieties. For **TAMRA-1** and **HiLyte-1**, α(2,3)Sia-terminated glycan–aldehyde was used. For **TAMRA-1'**, galactose-terminated glycan–aldehyde was alternatively used. In the next step, treatment with a variety of glyco-aldehyde probes afforded the desired heterogenous glycoalbumins (**TAMRA-2a**, **TAMRA-2b**, **TAMRA-2c**, **TAMRA-2c'**, **TAMRA-2d**, **TAMRA-2e**, **TAMRA-2f**, **HiLyte-2b**, **HiLyte-2c**, and **HiLyte-2e**). MALDI-TOF-MS analysis further confirms that the ratios of conjugated glycan moieties can be precisely controlled by adjusting the concentration of the two glycan–aldehyde probes.

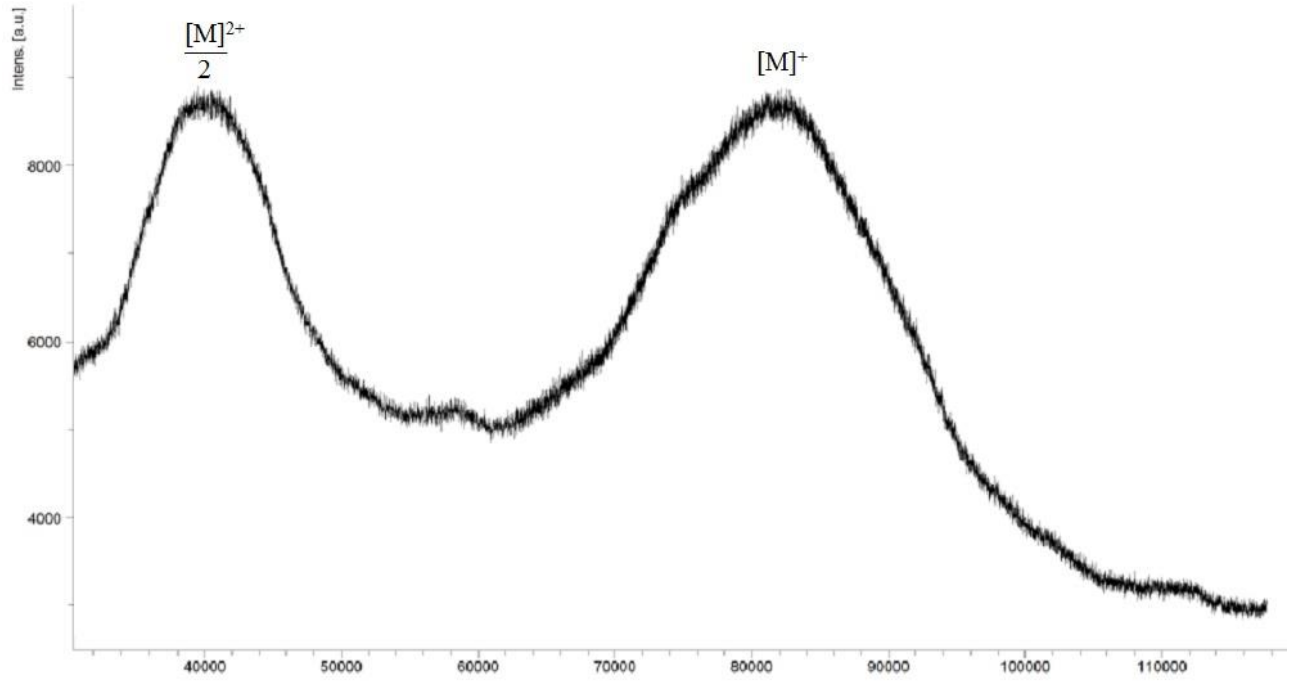


Scheme S1. Synthesis scheme for heterogeneous glycoalbumins **2a-f**. Synthesis of α(2,3)Sia- and Gal-terminated glycoalbumin **2c** is shown as an example.

Synthesis of intermediate **TAMRA-1**

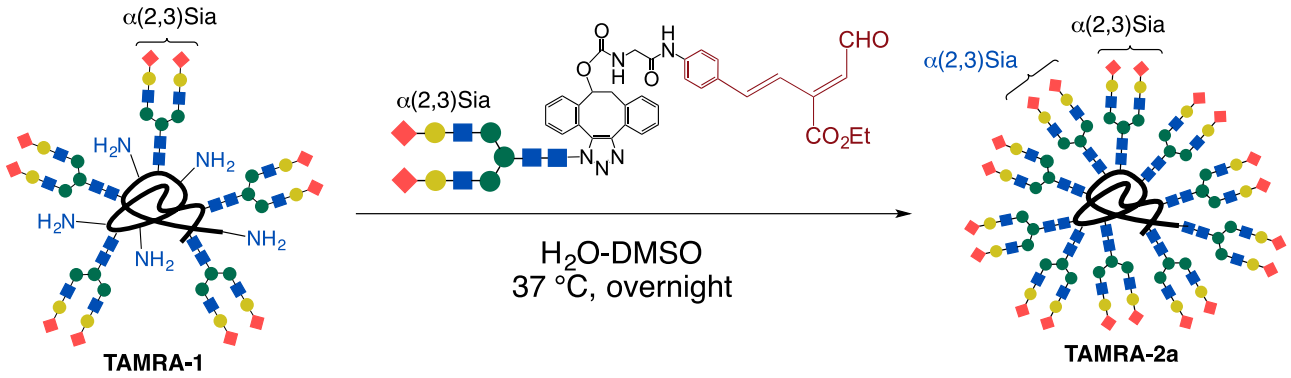


To **TAMRA-HSA** stock solution (96 μL, 12 nmol) was added water (307 μL), DMSO (96 μL), and then 3.8 mM stock solution of α(2,3)Sia-aldehyde in DMSO (144 nmol, 12 eq, 37 μL) under air. The mixture was incubated overnight at 37 °C to provide stock solution of **TAMRA-1**. A small amount of the reaction mixture (0.5 μL) was analyzed by MALDI-TOF-MS (positive mode), detecting the molecular weight of **TAMRA-1** at 82.5 kDa, which contains average number, 4.8 molecules of α(2,3)Sia-terminated disialoglycan per albumin.

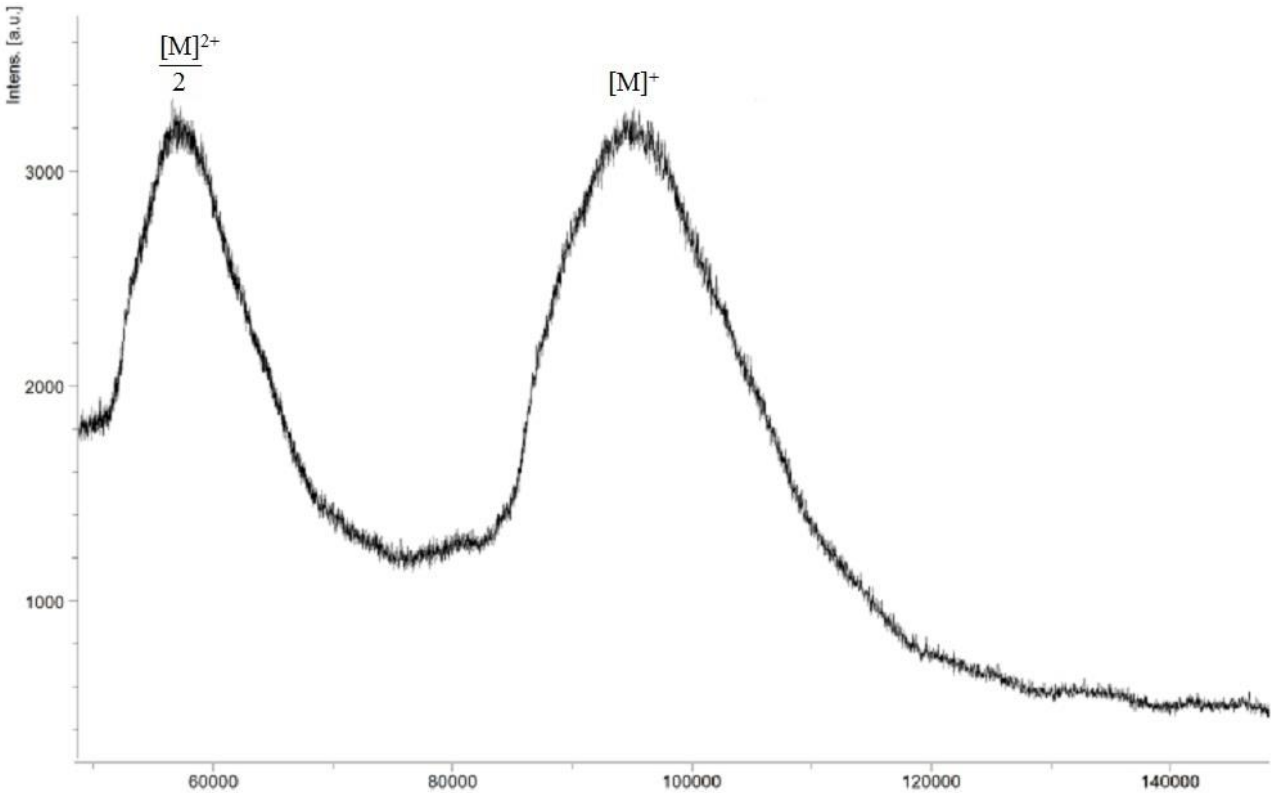


MALDI-TOF-MS of **TAMRA-1**

Preparation of homogeneous glycoalbumin **TAMRA-2a** with α(2,3)Sia

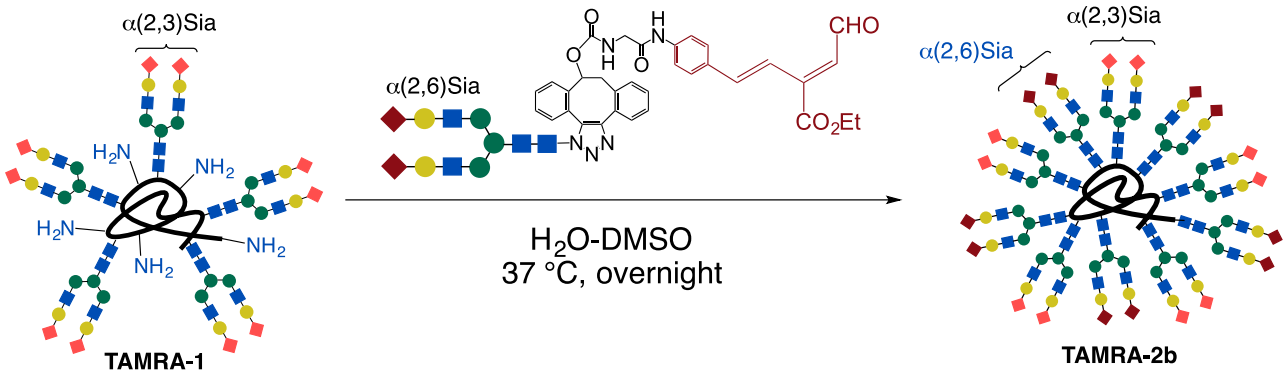


To **TAMRA-1** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of α(2,3)Sia-aldehyde in DMSO (17 nmol, 14 eq, 4.4 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of homogeneous glycoalbumin **TAMRA-2a**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2a** at 97.4 kDa, which therefore contained 4.9 molecules of glycan per albumin (total number of glycans introduced to albumin was 9.8).

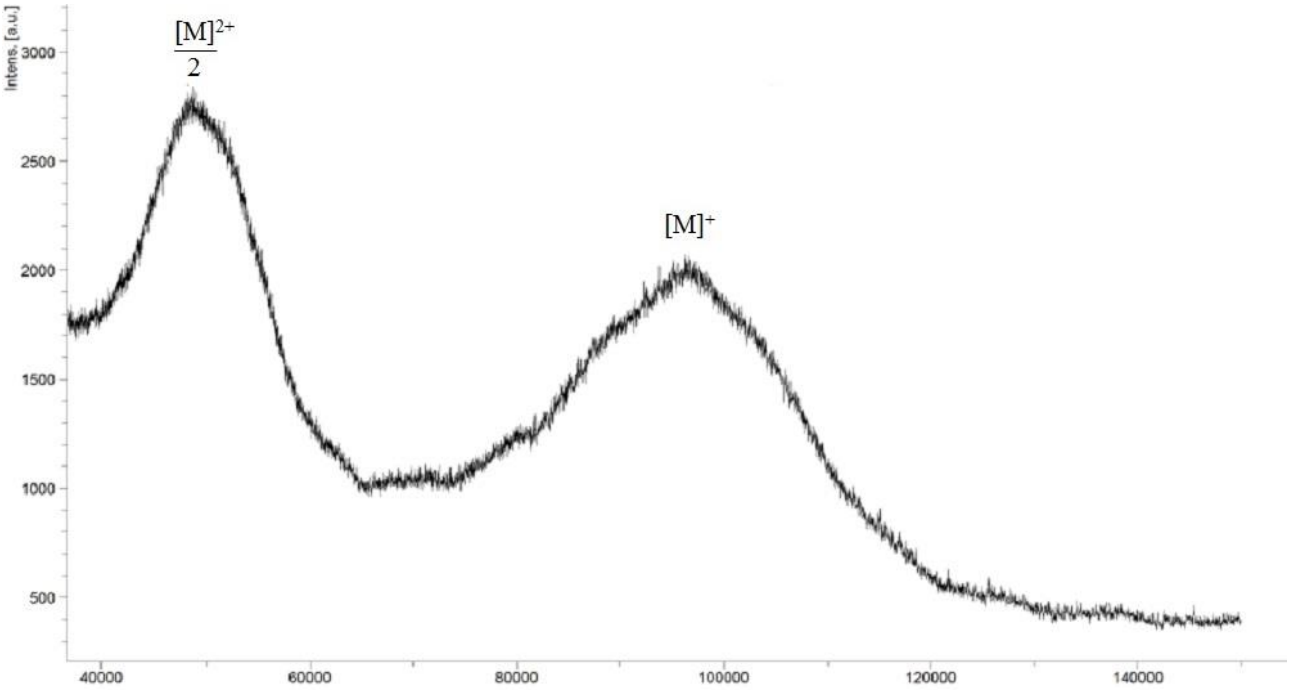


MALDI-TOF-MS of **TAMRA-2a**

Preparation of heterogeneous glycoalbumin **TAMRA-2b** with α(2,3)Sia and α(2,6)Sia

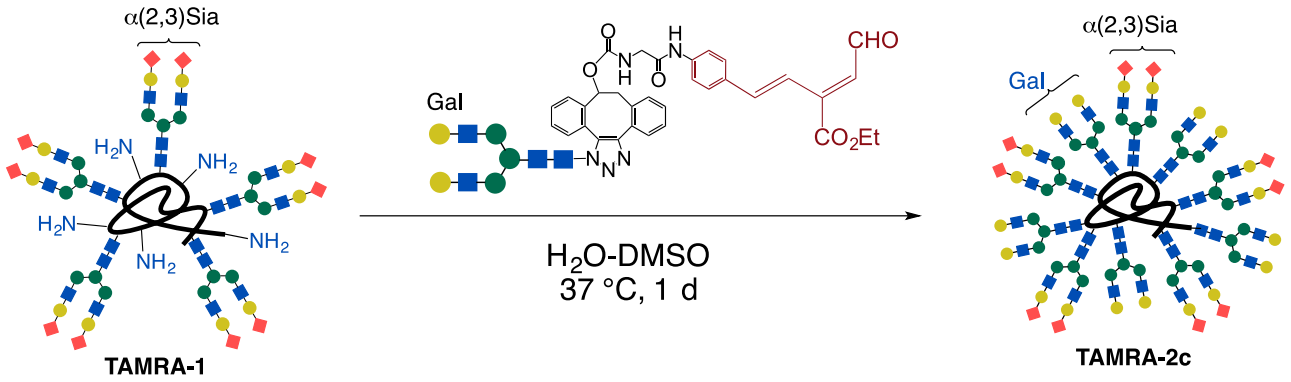


To **TAMRA-1** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of α(2,6)Sia-aldehyde in DMSO (19 nmol, 15 eq, 5.0 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2b**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2b** at 96.5 kDa, which therefore contained 4.6 molecules of glycan per albumin (total number of glycans introduced to albumin was 9.5).

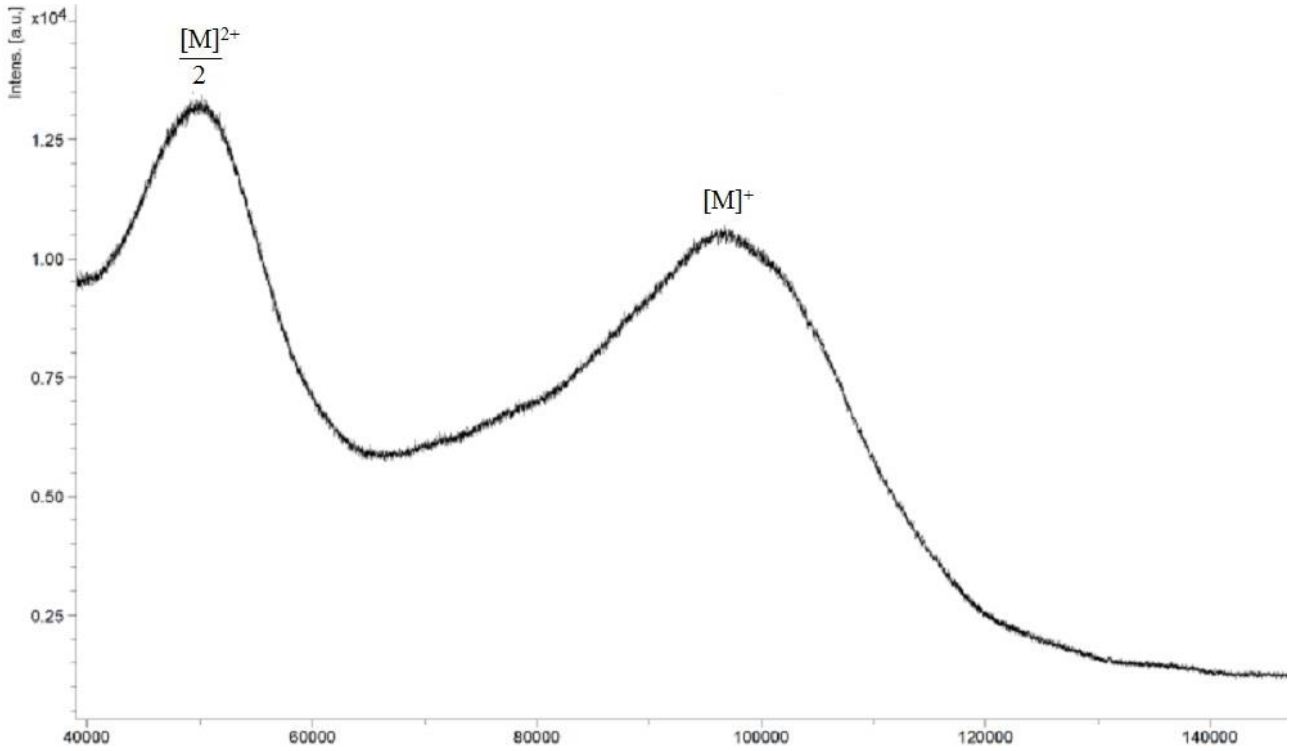


MALDI-TOF-MS of **TAMRA-2b**

Preparation of heterogeneous glycoalbumin **TAMRA-2c** with α(2,3)Sia and Gal

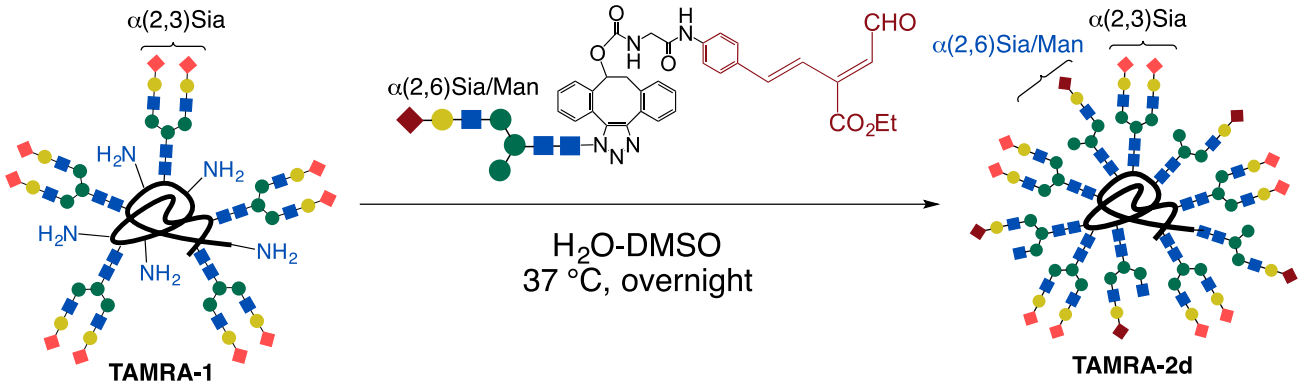


To **TAMRA-1** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of Gal-aldehyde in DMSO (21 nmol, 17 eq, 5.5 μL) in two portions under air. The mixture was incubated for 1 day at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2c**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2c** at 96.4 kDa, which therefore contained 5.7 molecules of glycan per albumin (total number of glycans introduced to albumin was 10.5).

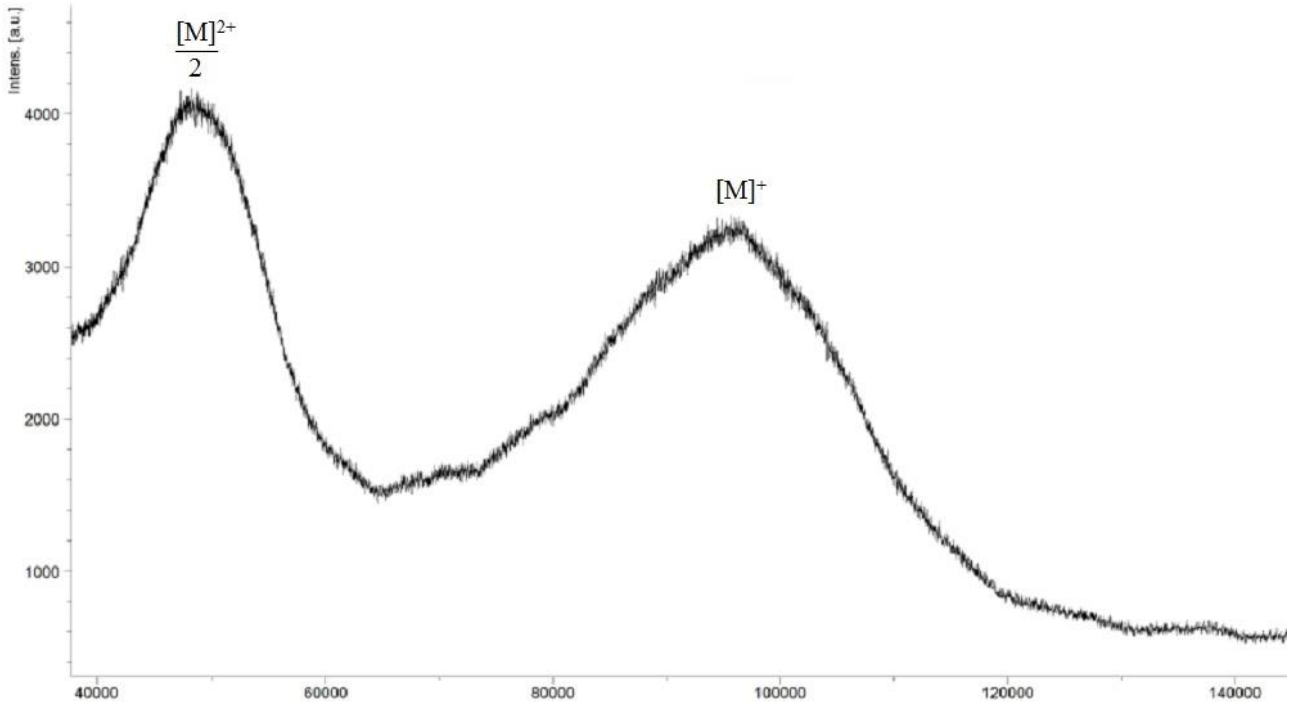


MALDI-TOF-MS of **TAMRA-2c**

Preparation of heterogeneous glycoalbumin **TAMRA-2d** with α(2,3)Sia and α(2,6)Sia/Man

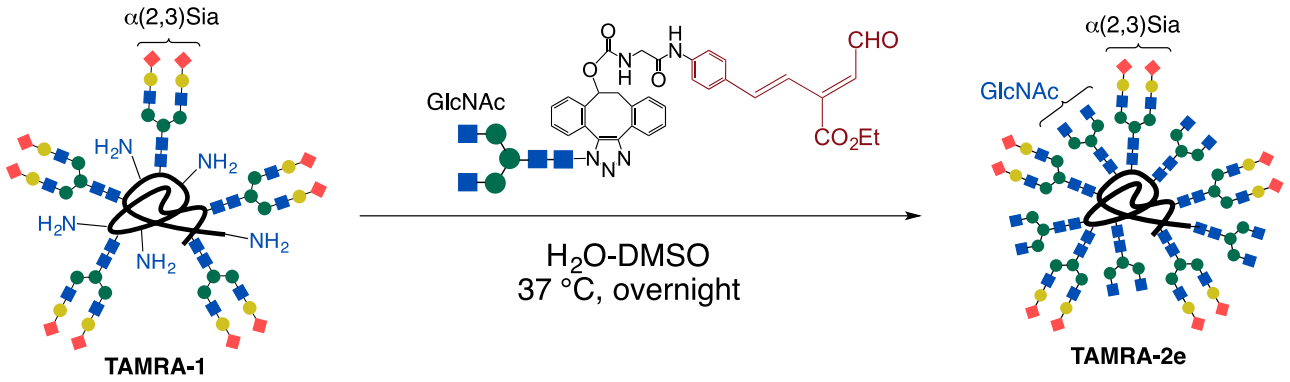


To **TAMRA-1** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of α(2,6)Sia/Man-aldehyde in DMSO (15 nmol, 12 eq, 3.9 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2d**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2d** at 96.0 kDa, which therefore contained 5.7 molecules of glycan per albumin (total number of glycans introduced to albumin was 10.6).

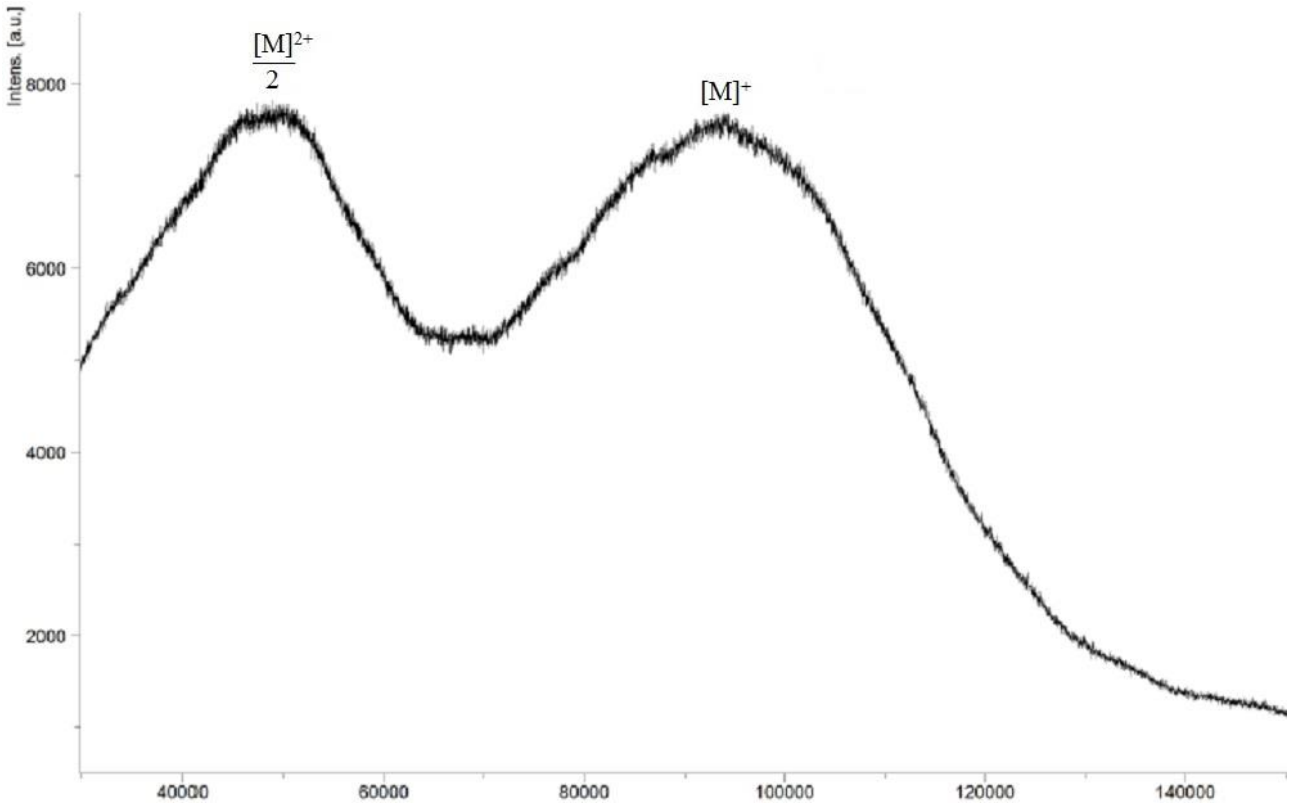


MALDI-TOF-MS of **TAMRA-2d**

Preparation of heterogeneous glycoalbumin **TAMRA-2e** with α(2,3)Sia and GlcNAc

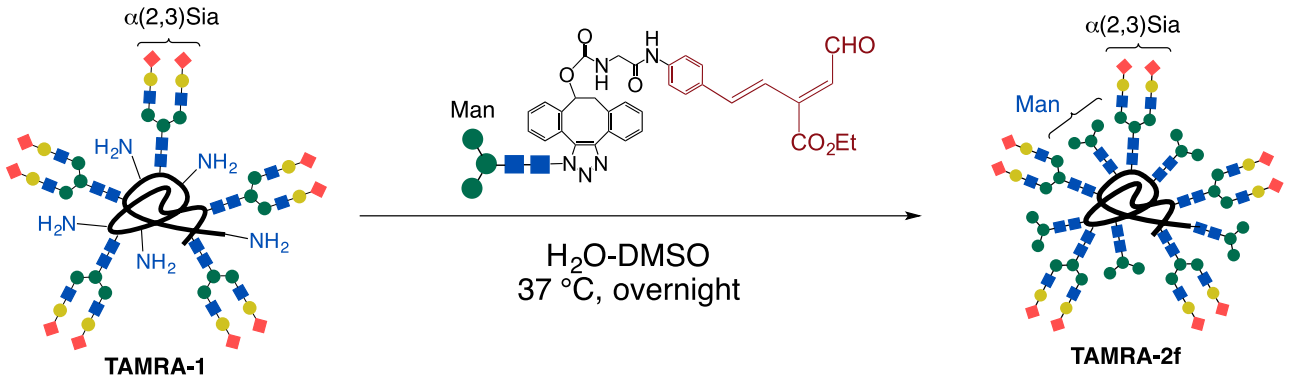


To **TAMRA-1** stock solution (62 μL, 1.5 nmol) was added 3.8 mM stock solution of GlcNAc-aldehyde in DMSO (20 nmol, 14 eq, 5.4 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2e**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2e** at 92.8 kDa, which therefore contained 4.9 molecules of glycan per albumin (total number of glycans introduced to albumin was 9.7).

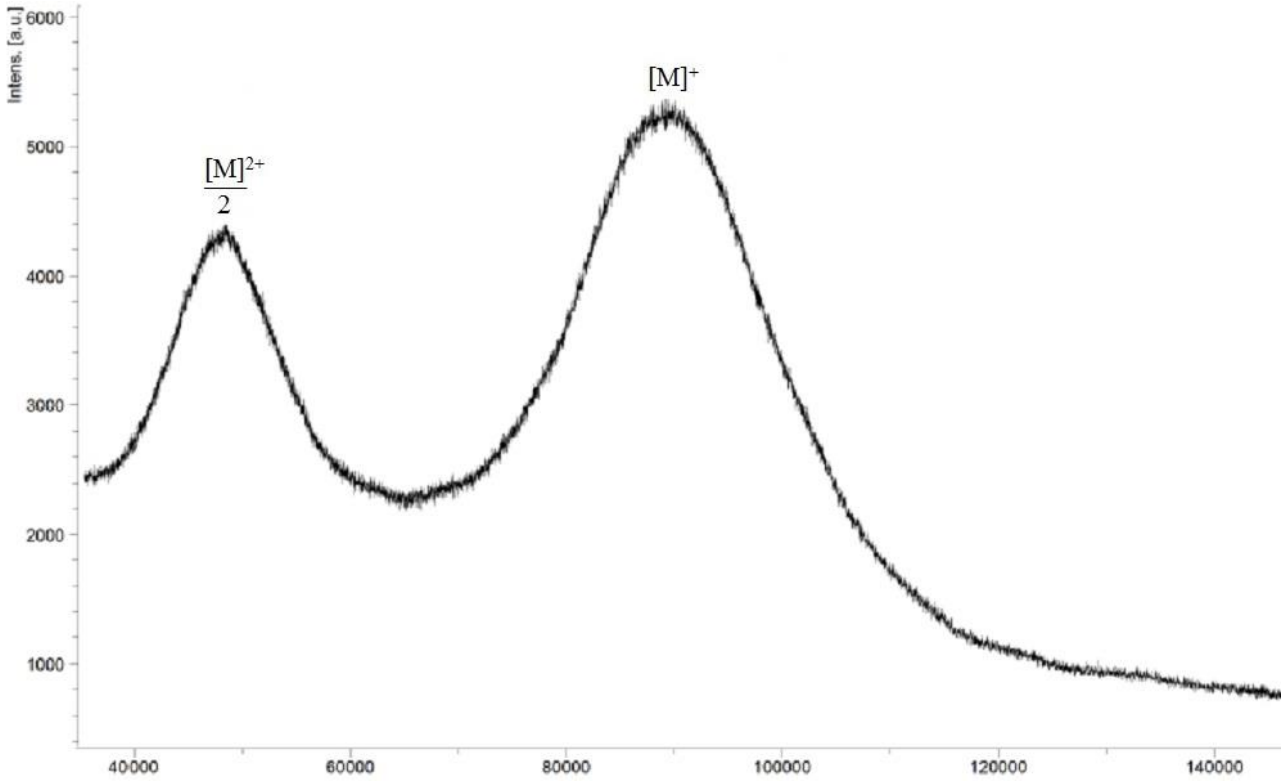


MALDI-TOF-MS of **TAMRA-2e**

Preparation of heterogeneous glycoalbumin **TAMRA-2f** with α(2,3)Sia and Man

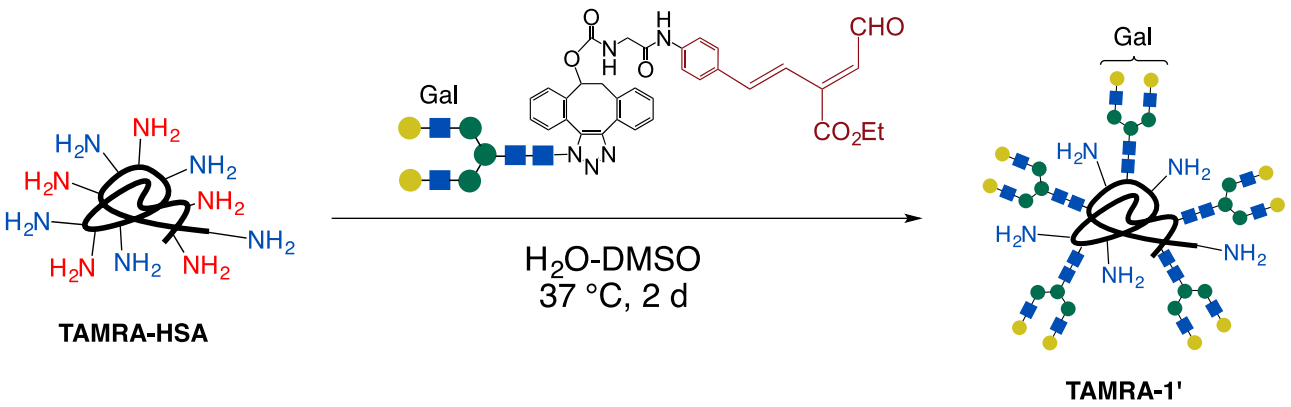


To **TAMRA-1** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of Man-aldehyde in DMSO (15 nmol, 12 eq, 4.0 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2f**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2f** at 90.9 kDa, which therefore contained 4.9 molecules of glycan per albumin (total number of glycans introduced to albumin was 9.7).

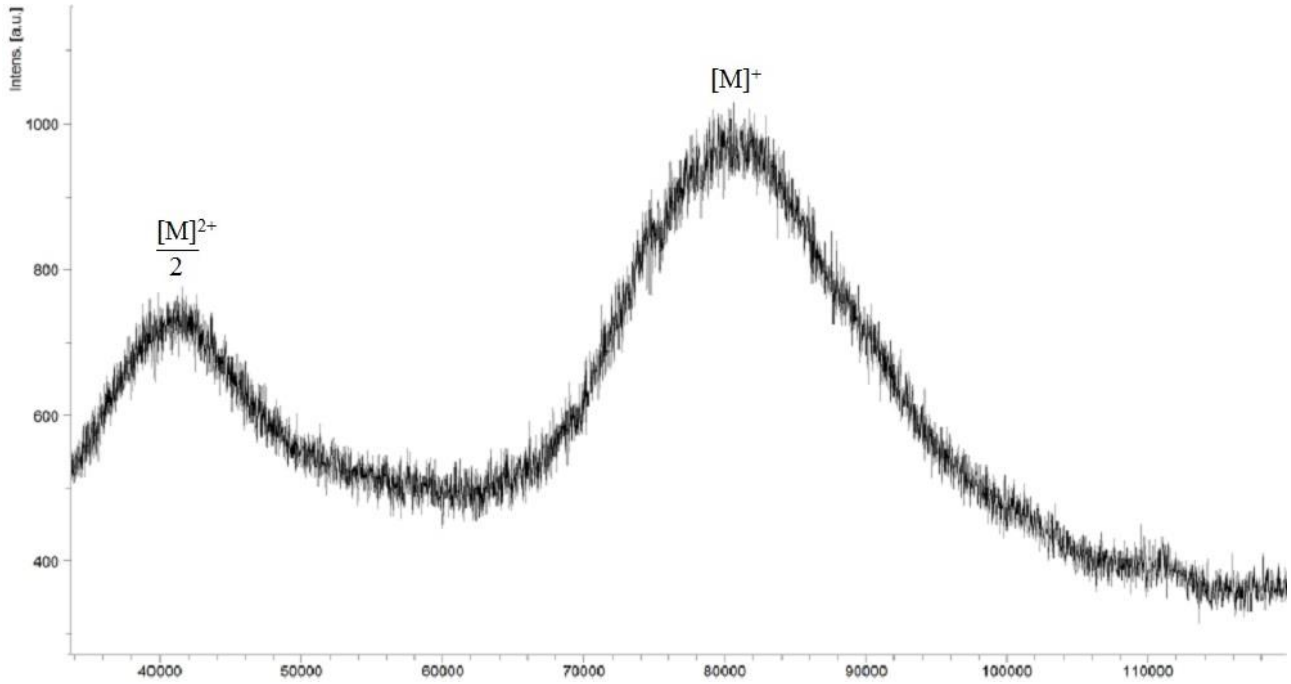


MALDI-TOF-MS of **TAMRA-2f**

Synthesis of intermediate **TAMRA-1'**

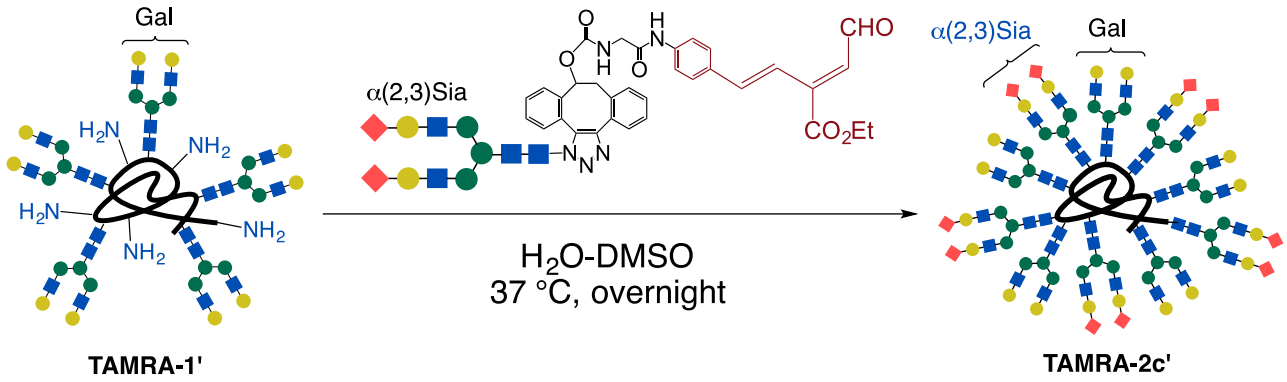


To **TAMRA-HSA** stock solution (32 μL, 4.0 nmol) was added water (102 μL), DMSO (32 μL), and then 3.8 mM stock solution of Gal-aldehyde in DMSO (67 nmol, 18 eq, 17 μL) in three portions under air. The mixture was incubated for 2 d at 37 °C to provide stock solution of **TAMRA-1'**. A small amount of the reaction mixture (0.5 μL) was analyzed by MALDI-TOF-MS (positive mode), detecting the molecular weight of **TAMRA-1'** at 80.9 kDa, which contains average number, 5.3 molecules of Gal-terminated glycan per albumin.

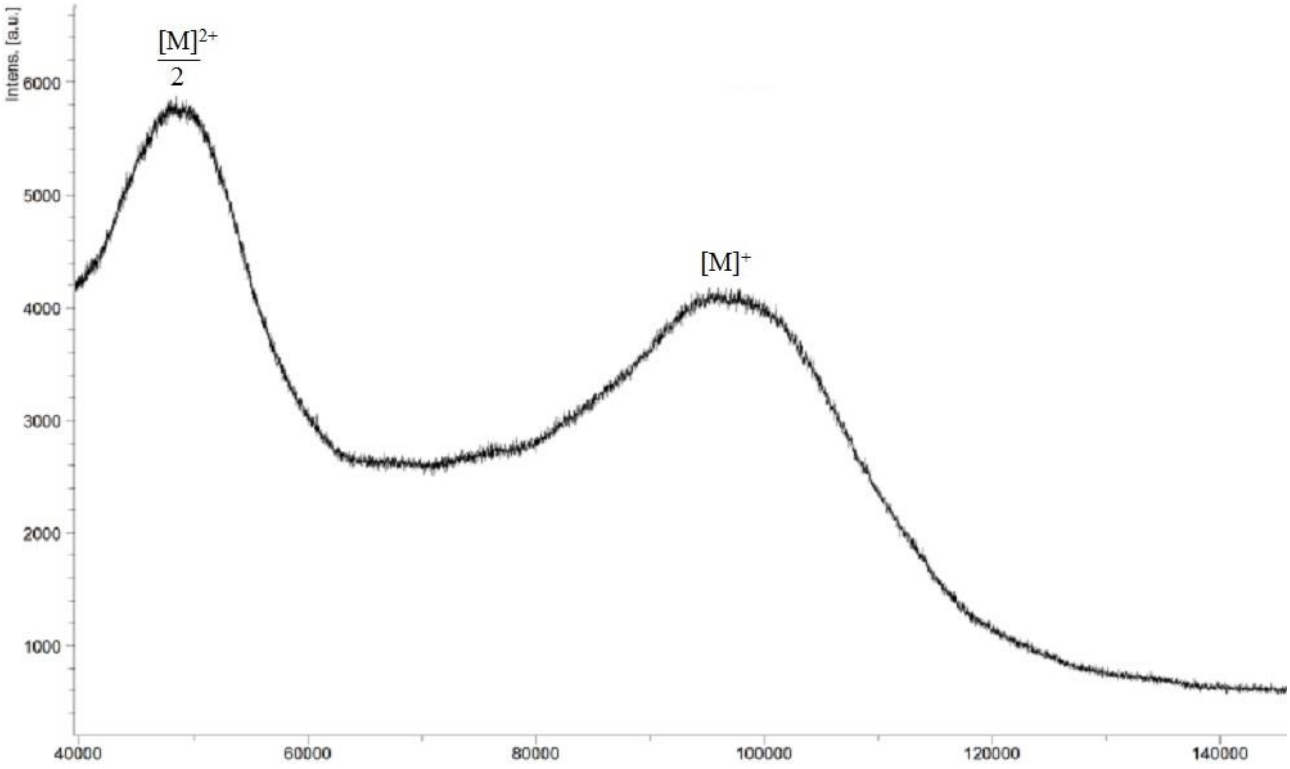


MALDI-TOF-MS of **TAMRA-1'**

Preparation of heterogeneous glycoalbumin **TAMRA-2c'** with Gal and α(2,3)Sia

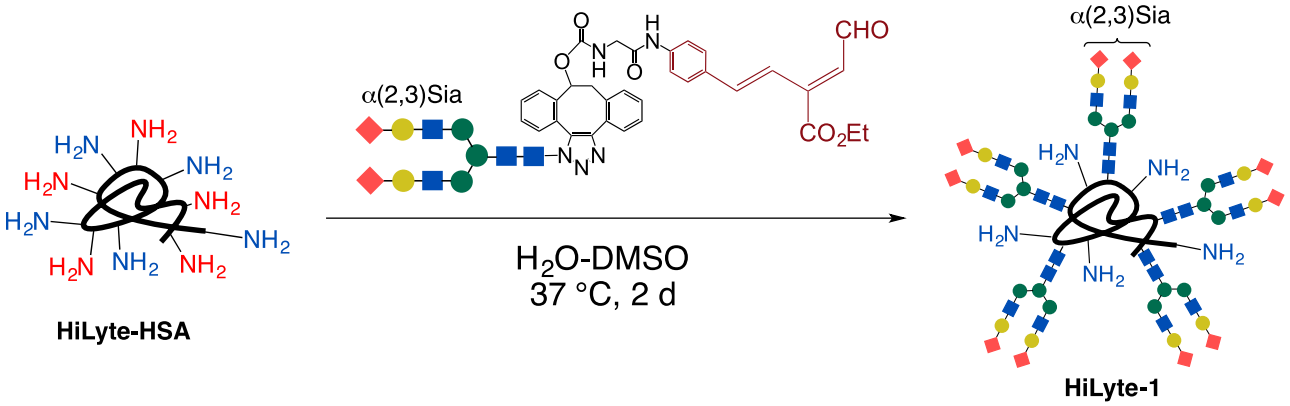


To **TAMRA-1'** stock solution (52 μL, 1.25 nmol) was added 3.8 mM stock solution of α(2,3)Sia-aldehyde in DMSO (17 nmol, 14 eq, 4.4 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **TAMRA-2c'**. MALDI-TOF-MS (positive mode) detected the molecular weight of **TAMRA-2c'** at 97.5 kDa, which therefore contained 5.5 molecules of glycan per albumin (total number of glycans introduced to albumin was 10.8).

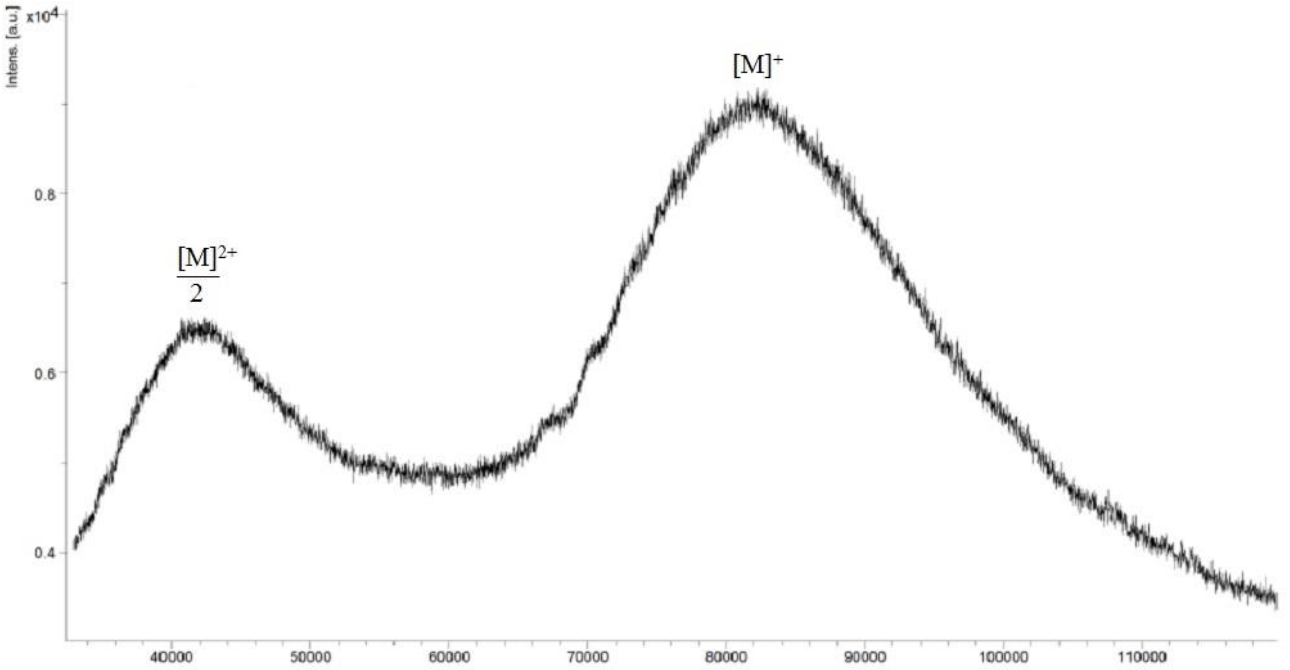


MALDI-TOF-MS of **TAMRA-2c'**

Synthesis of intermediate **HiLyte-1**

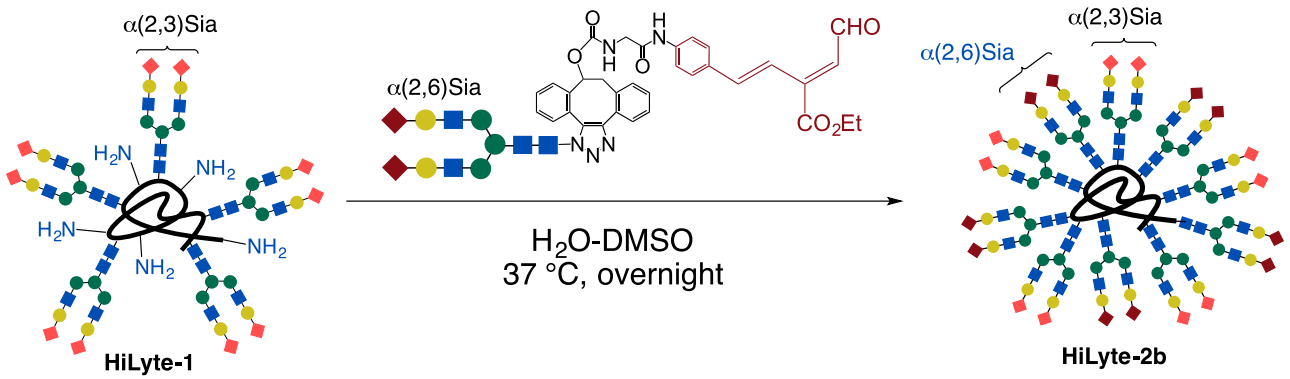


To **HiLyte-HSA** stock solution (175 μL, 10 nmol) was added water (175 μL), DMSO (88 μL), and then 3.8 mM stock solution of α(2,3)Sia-aldehyde in DMSO (179 nmol, 18 eq, 47 μL) in three portions under air. The mixture was incubated for 2 days at 37 °C to provide stock solution of **HiLyte-1**. A small amount of the reaction mixture (0.5 μL) was analyzed by MALDI-TOF-MS (positive mode), detecting the molecular weight of **HiLyte-1** at 83.9 kDa, which contains average number, 4.5 molecules of α(2,3)Sia-terminated disialoglycan per albumin.

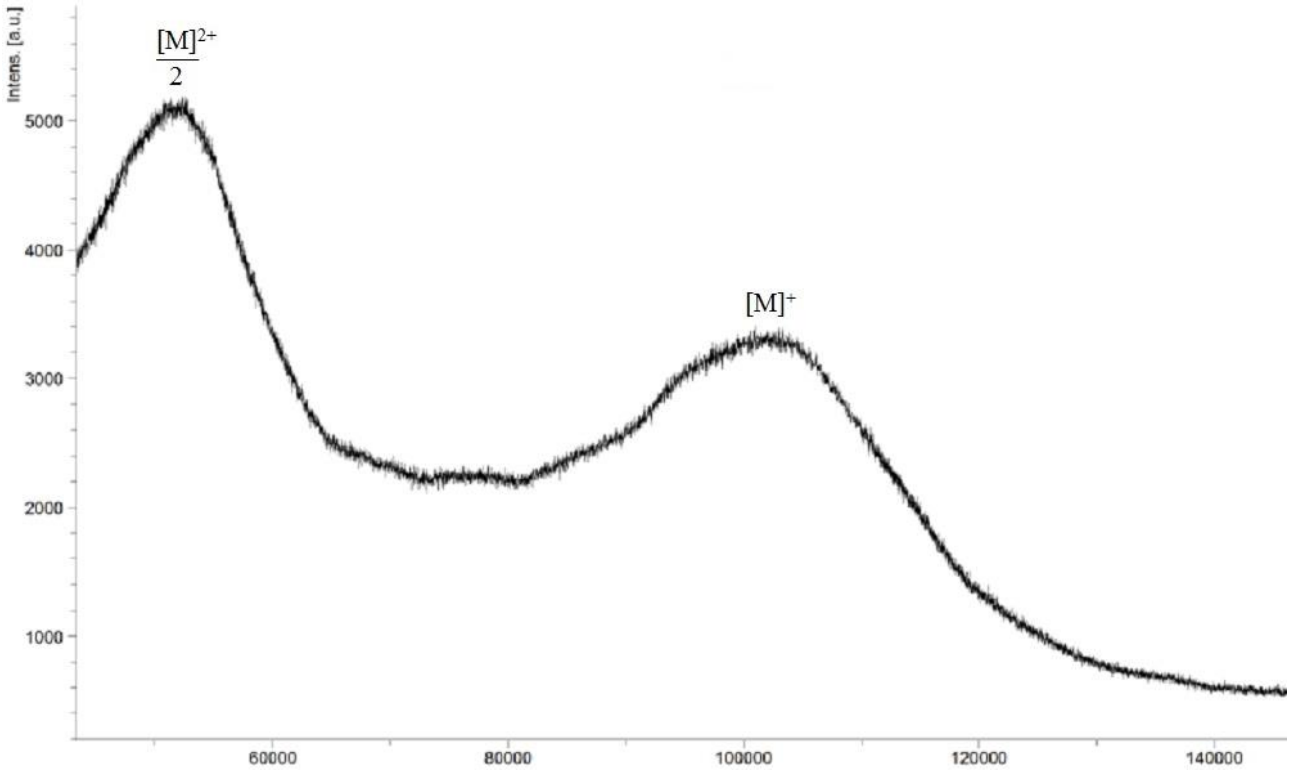


MALDI-TOF-MS of **HiLyte-1**

Preparation of heterogeneous glycoalbumin **HiLyte-2b** with α(2,3)Sia and α(2,6)Sia

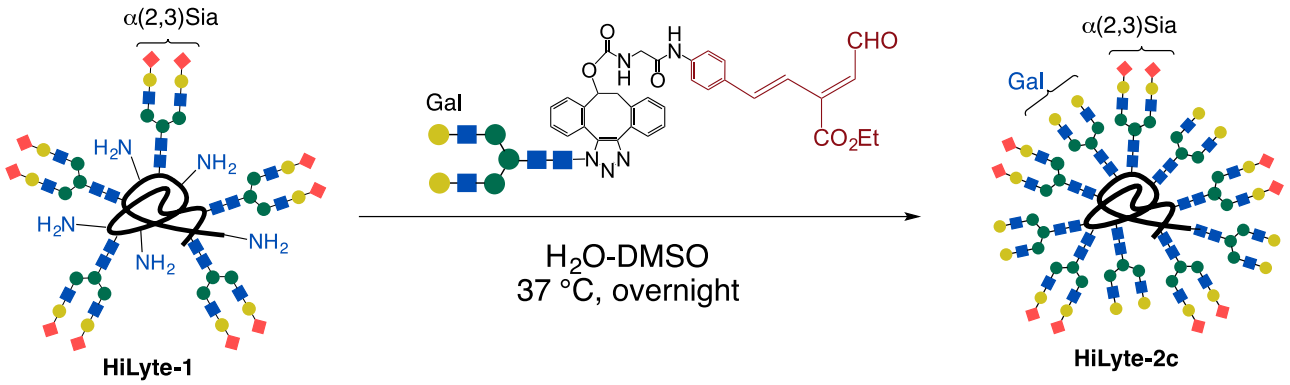


To **HiLyte-1** stock solution (43 μL, 1 nmol) was added 3.8 mM stock solution of α(2,6)Sia-aldehyde in DMSO (9 nmol, 9 eq, 2.4 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **HiLyte-2b**. MALDI-TOF-MS (positive mode) detected the molecular weight of **HiLyte-2b** at 101.7 kDa, which therefore contained 5.9 molecules of glycan per albumin (total number of glycans introduced to albumin was 10.4).

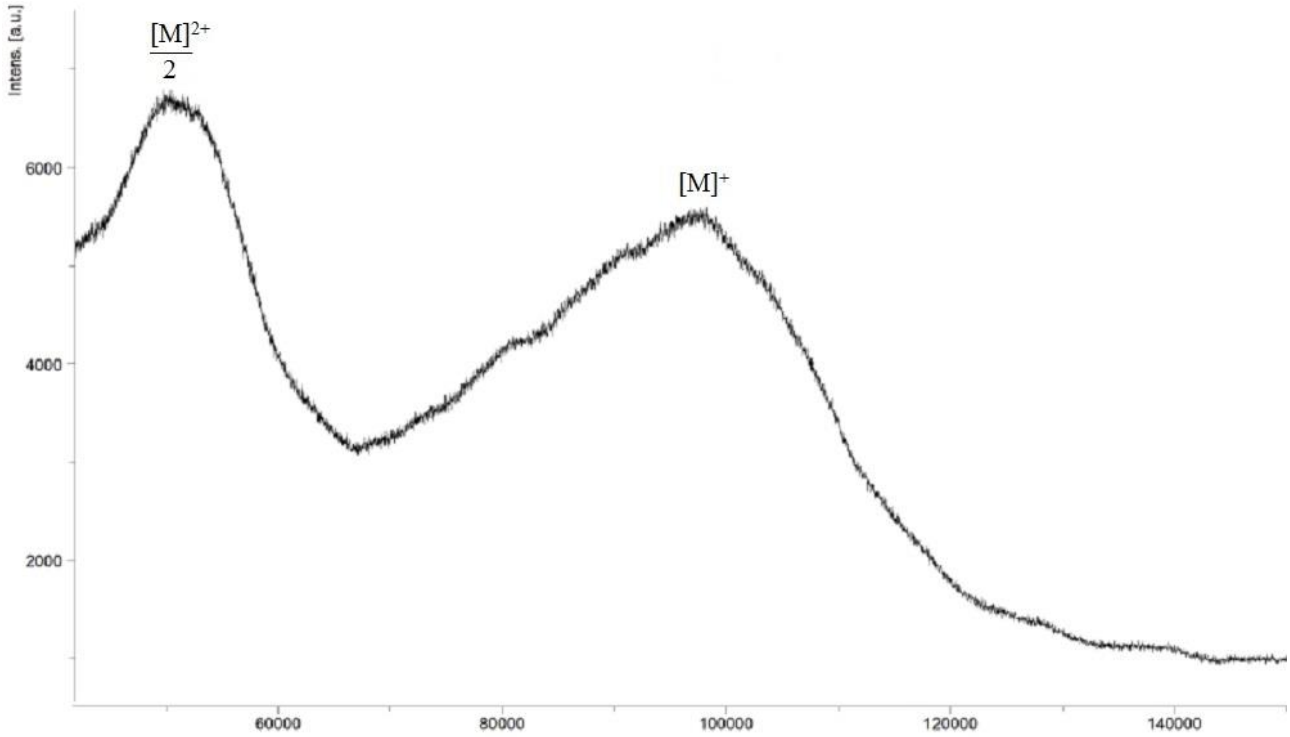


MALDI-TOF-MS of **HiLyte-2b**

Preparation of heterogeneous glycoalbumin **HiLyte-2c** with α(2,3)Sia and Gal

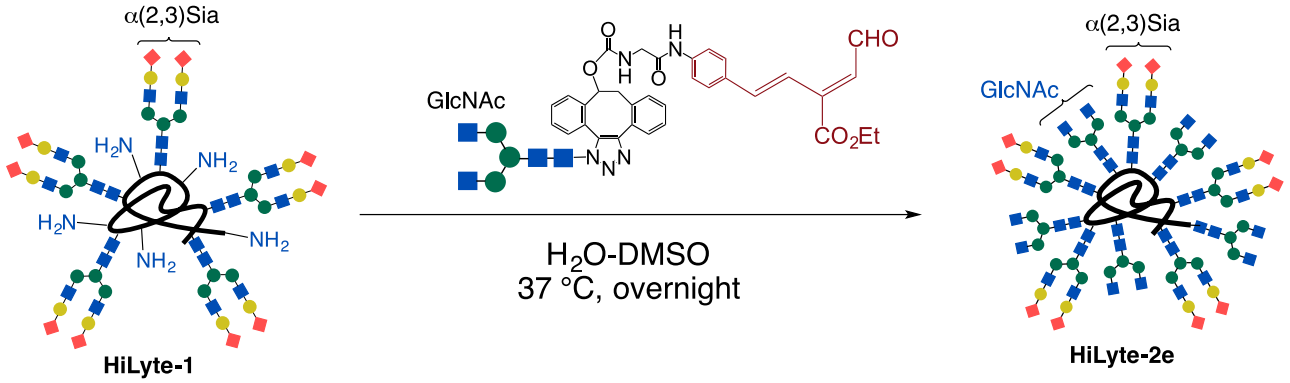


To **HiLyte-1** stock solution (43 μL, 1 nmol) was added 3.8 mM stock solution of Gal-aldehyde in DMSO (11 nmol, 11 eq, 2.9 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **HiLyte-2c**. MALDI-TOF-MS (positive mode) detected the molecular weight of **HiLyte-2c** at 98.8 kDa, which therefore contained 6.1 molecules of glycan per albumin (total number of glycans introduced to albumin was 10.6).

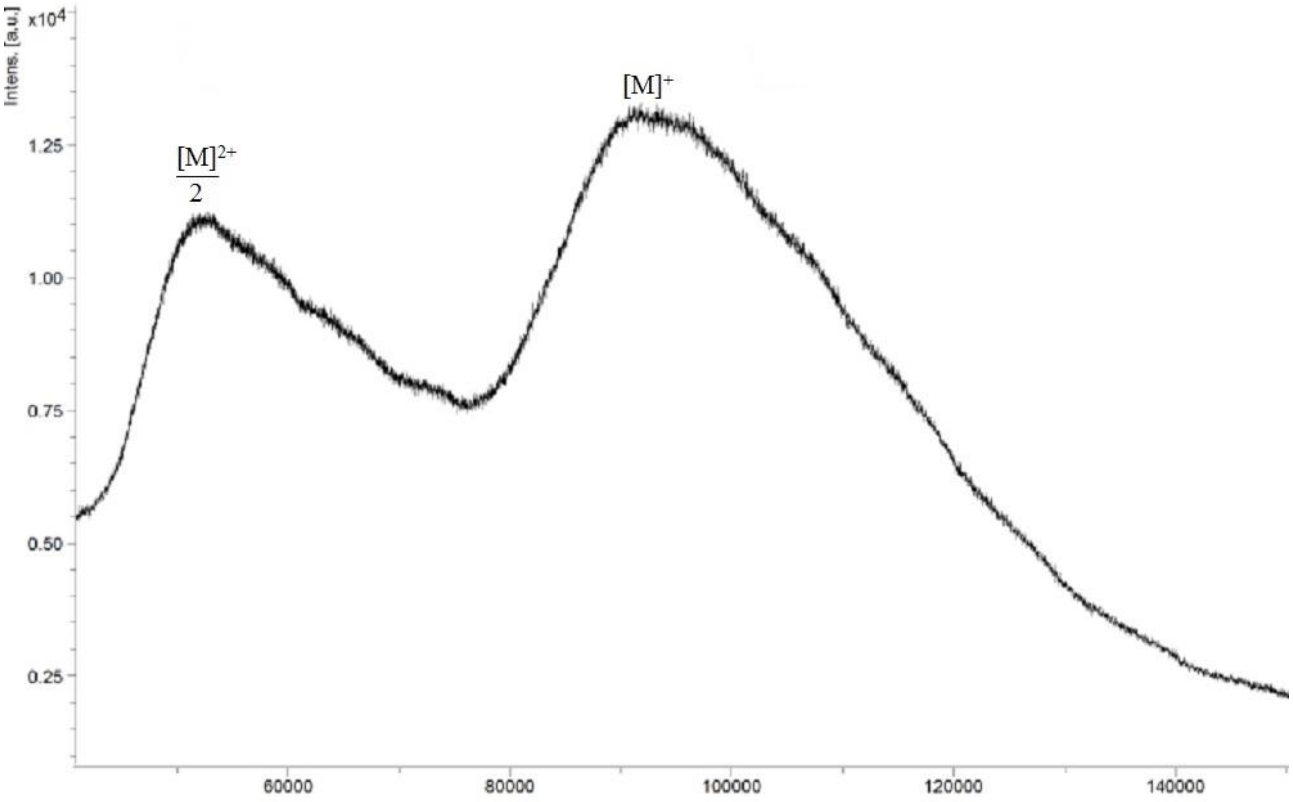


MALDI-TOF-MS of **HiLyte-2c**

Preparation of heterogeneous glycoalbumin **HiLyte-2e** with α(2,3)Sia and GlcNAc



To **HiLyte-1** stock solution (43 μL, 1 nmol) was added 3.8 mM stock solution of GlcNAc-aldehyde in DMSO (11 nmol, 11 eq, 2.9 μL) under air. The mixture was incubated overnight at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 10 μM solution of heterogeneous glycoalbumin **HiLyte-2e**. MALDI-TOF-MS (positive mode) detected the molecular weight of **HiLyte-2e** at 94.0 kDa, which therefore contained 4.7 molecules of glycan per albumin (total number of glycans introduced to albumin was 9.2).



MALDI-TOF-MS of **HiLyte-2e**

1.4 Glycoalbumin Ligation Site Determination

To determine the lysine residues that are favored for conjugation during the formation of the heterogeneous glycoalbumins used in this study, a mass spectrometry-based study was performed using two model glycoalbumins. **5Gal-HSA** is a glycoalbumin conjugated with approximately 5 galactose terminated glycans (to mimic the intermediate after the first coupling step), while **10Gal-HSA** is a glycoalbumin conjugated with approximately 10 galactose terminated glycans (to mimic the product after both coupling steps). Using two separate mass spectrometry techniques (LC-MS/MS and MALDI-TOF/MS), the results were consolidated to reveal 10 lysine residues that are likely ligation sites for glycoalbumins **2a-f**.

1.4.1 Summary

In summary, 10 lysine residues (Lys73, Lys136, Lys195, Lys199, Lys351, Lys402, Lys439, Lys525, Lys536, Lys541) were identified as preferential ligation sites with the glycan–aldehyde probes used in this study. Summarized in Figure S2, these lysine residues are displayed in red.

```
  1 DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFA
 51 KTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNE
101 CFLQHKDDNPNLPRLVRPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFY
151 APELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDEGKASSAKQRLKC
201 ASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDL
251 LECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPA
301 DLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRLA
351 KTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGE
401 YKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAE
451 DYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPK
501 EFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDD
551 FAAFVEKCCKADDKETCFAEEGKKLVAASQAALGL
```

Figure S2. Summary of specific lysine residues in human serum albumin prone to ligation with glycan–aldehyde probes used in this study. This data was collected via consolidation from LC-MS/MS and MALDI-TOF/MS analyses.

By comparing the abundance ratios from the LC-MS/MS experiments and the signal intensities from the MALDI-TOF/MS experiments, it is possible to estimate which of the aforementioned lysines are preferentially labeled in **5Gal-HSA**. In theory, these lysines would represent the amino acids that react the fastest. These results are summarized in Table S1.

Table S1. Summary of lysine ligation sites

	5Gal-HSA	10Gal-HSA
Lys73		✔
Lys136		✔
Lys195	✔	✔
Lys199	✔	✔
Lys351		✔
Lys402		✔
Lys439		✔
Lys525	✔	✔
Lys536	✔	✔
Lys541	✔	✔

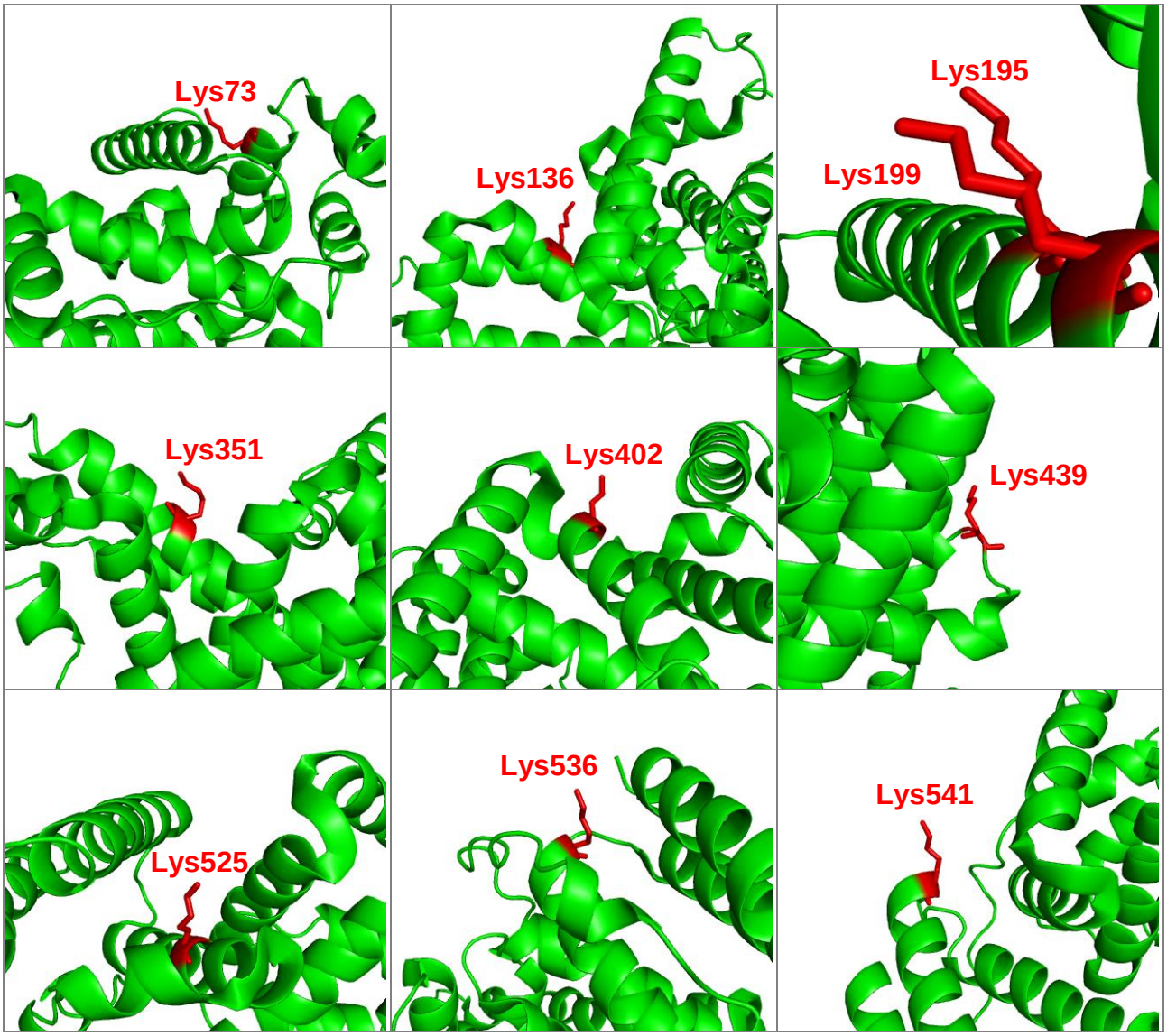
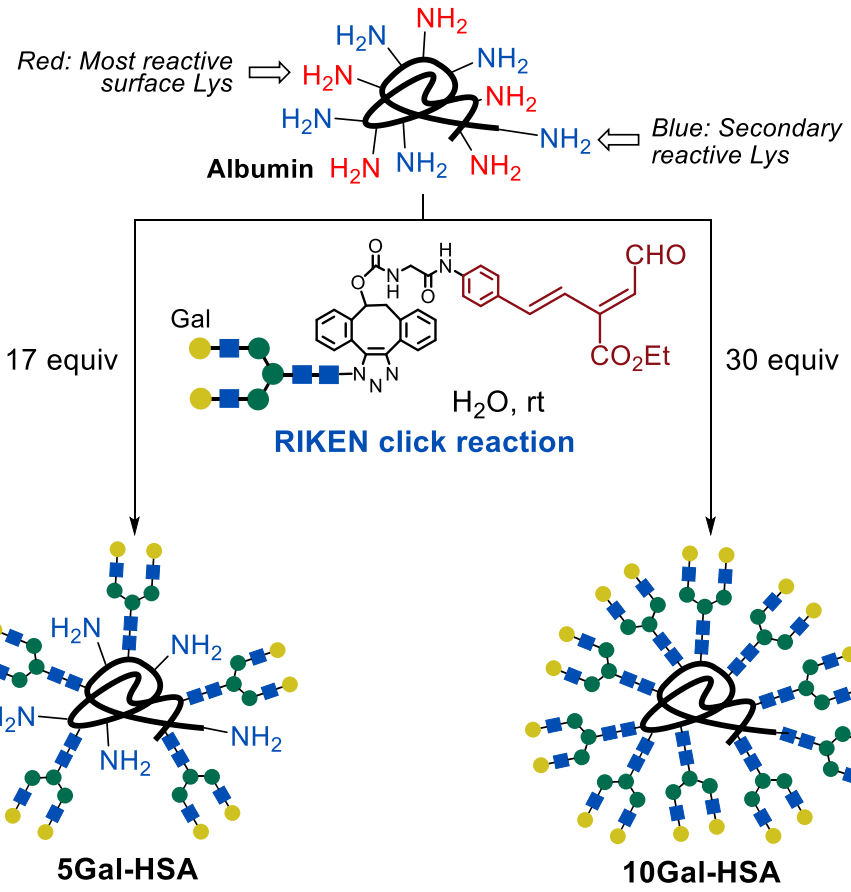


Figure S3. Structure of HSA (PDB - 2BXI) with highlighted (in red) lysine residues that are proposed to be conjugated with glycans in the glycoalbumins used for this study. In general, all lysines were found to be sufficiently exposed to the surface.

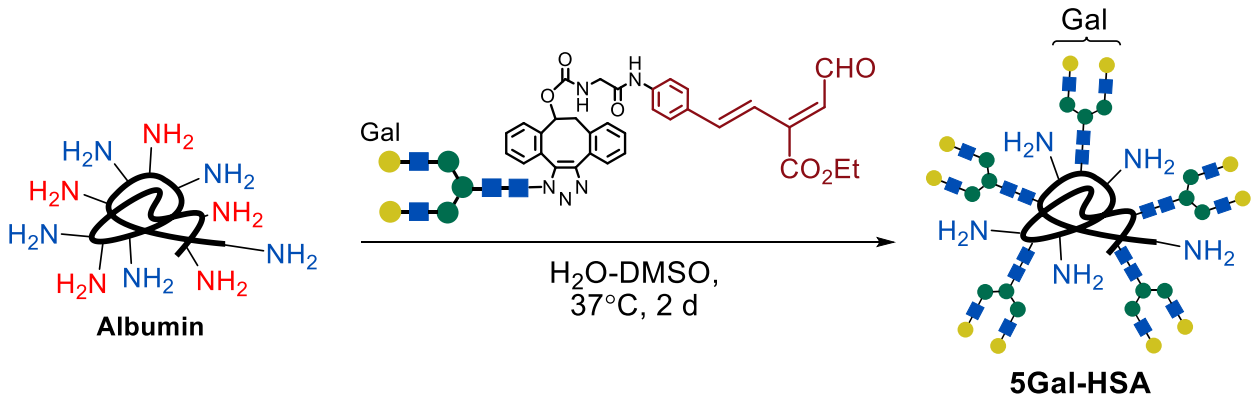
1.4.2 Preparation of Glycoalbumins **5Gal-HSA**, **10Gal-HSA**

In the general synthetic scheme of **5Gal-HSA** and **10Gal-HSA** (Scheme S2), the key determining factor was the reaction concentration of the galactose-terminated glycan–aldehyde. Human serum albumin (HSA) was either treated with 17 equivalents or 30 equivalents of the galactose-terminated glycan–aldehyde, which leads to either **5Gal-HSA** or **10Gal-HSA**, respectively. MALDI-TOF-MS analysis can then be performed to confirm the number of conjugated glycan moieties.

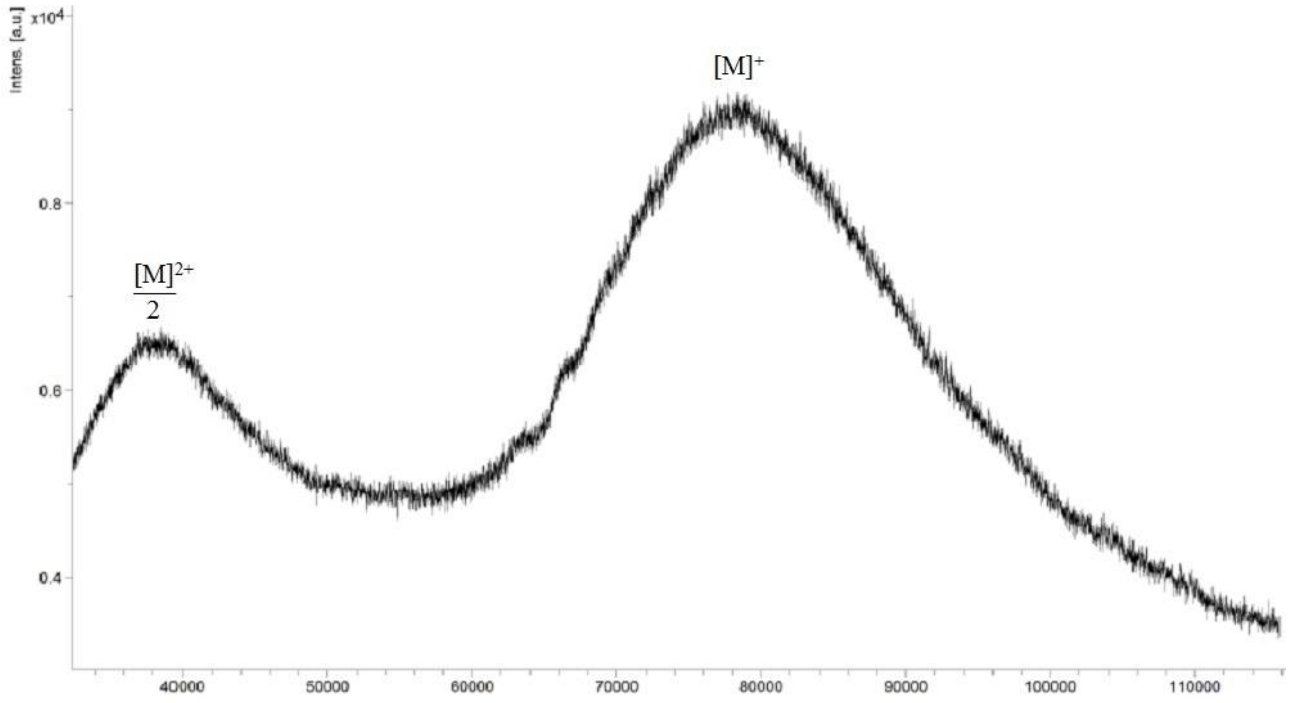


Scheme S2. Synthesis scheme for homogenous Gal-terminated glycoalbumins **5Gal-HSA** and **10Gal-HSA**.

Preparation of homogeneous glycoalbumin **5Gal-HSA** with Gal

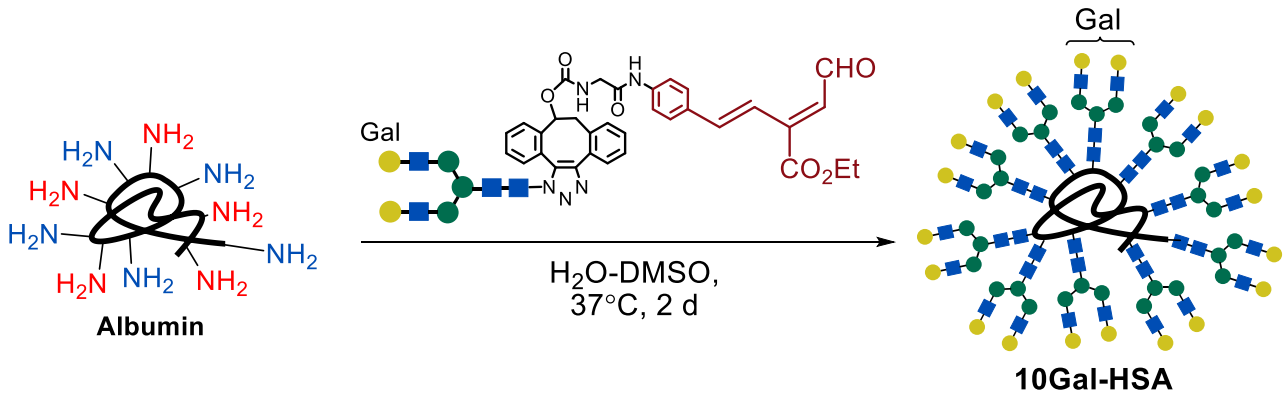


To **HSA** stock solution (36 μL, 6.0 nmol) was added water (288 μL), DMSO (96 μL), and then 3.8 mM stock solution of Gal-aldehyde in DMSO (102 nmol, 17 eq, 27 μL) under air. The mixture was incubated for 2 d at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 50 μM solution of homogeneous glycoalbumin **5Gal-HSA**. MALDI-TOF-MS (positive mode) detected the molecular weight of **5Gal-HSA** at 79.2 kDa, which contains average number, 5.1 molecules of Gal-terminated glycan per albumin.

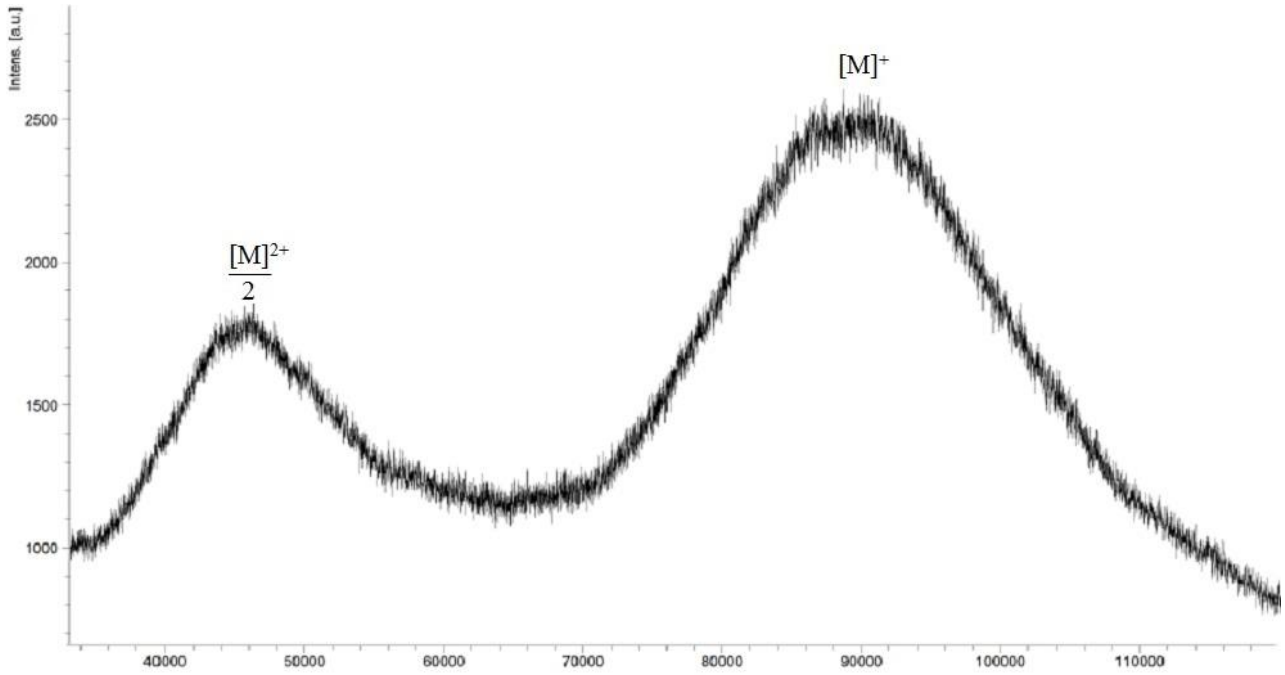


MALDI-TOF-MS of **5Gal-HSA**

Preparation of homogeneous glycoalbumin **10Gal-HSA** with Gal



To **HSA** stock solution (64 μL, 4.0 nmol) was added water (192 μL), DMSO (64 μL), and then 3.8 mM stock solution of Gal-aldehyde in DMSO (120 nmol, 30 eq, 32 μL) under air. The mixture was incubated for 2 d at 37 °C. The resulting solution was diluted with water and centrifuged through Amicon 10K® at 15,000 rpm for 10 min, and further washed with water three times to filter off any small molecules. The insoluble byproducts were removed by filtering with Durapore PVDF 0.45 μm® and diluted with water to give 50 μM solution of homogeneous glycoalbumin **10Gal-HSA**. MALDI-TOF-MS (positive mode) detected the molecular weight of **10Gal-HSA** at 91.0 kDa, which contains average number, 9.8 molecules of Gal-terminated asialoglycan per albumin.



MALDI-TOF-MS of **10Gal-HSA**

1.4.3 LC-MS/MS Analysis

For the LC-MS/MS study, three different protein samples were digested and analyzed (native HSA, **5Gal-HSA**, and **10Gal-HSA**). Endoproteinase Asp-N was employed for enzymatic digestion, which acts by selectively cleaving peptide bonds on the N-terminal side of aspartic acid (Asp, D) and glutamic acid (Glu, E) residues.

Analysis was primarily performed by measuring and comparing the abundance ratios of found peptide fragments between the native HSA with either **5Gal-HSA** or **10Gal-HSA**. To establish that the concentration of analyzed protein samples were roughly equivalent, the overall abundance ratio of **5Gal-HSA**/HSA was measured to be 0.995, while that of **10Gal-HSA**/HSA was measured to be 0.874.

A summary of suspected lysine residues (shown in orange) that may act as ligation sites are shown in Figure S4. Analyzed peptide fragments which this data was obtained from are discussed and examined in Figures S5-8.

Detailed MS chromatograms, MS/MS spectra, and summary of detected fragment matches can be found in Figures S9-20.

```
  1 DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFA
 51 KTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNE
101 CFLQHKDDNPNLPRLVRPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFY
151 APELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDEGKASSAKQRLKC
201 ASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDL
251 LECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPA
301 DLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRLA
351 KTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGE
401 YKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAE
451 DYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPK
501 EFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDD
551 FAAFVEKCCKADDKETCFAEEGKKLVAASQAALGL
```

Figure S4. Summary of suspected lysine residues in human serum albumin that are prone to ligation with glycan–aldehyde probes obtained via LC-MS/MS methods.

Figure S5 shows albumin residues 1-100, along with the corresponding peptide fragments L1-L7 that have notable abundance ratio changes throughout the LC-MS/MS studies. Lysine residues of interest were identified at positions 4, 73, and 93.

- For Lys4, peptide fragment L1 designates it as a fair/good site of ligation. To deduce whether Lys4 or Lys12 is the ligation site of interest, fragment L2 was then identified for its generally poor abundance ratio, thereby affirming Lys4 as the ligation site.
- For Lys73, peptide fragments L3 and L4 designates it as a fair/good site of ligation.
- For Lys 93, peptide fragments L5-L7 designates it as a good site of ligation.

```
 1 DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFA
   DAHKSEVAHRFK------------------------------------ L1
   -----EVAHRFK------------------------------------ L2
   ------------------------------------------------ L3
   ------------------------------------------------ L4
   ------------------------------------------------ L5
   ------------------------------------------------ L6
   ------------------------------------------------ L7

51 KTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNE
   ------------------------------------------------ L1
   ------------------------------------------------ L2
   ----------------------DKLCTVATLRETYGEMA---------- L3
   ----------------------DKLCTVATLRETYG------------- L4
   -------------------------------ETYGEMADCCAKQEPERN- L5
   -------------------------------ETYGEMADCCAKQ------ L6
   ------------------------------------DCCAKQEP---- L7
```

	Abundance Ratios			
	5Gal-HSA/HSA		**10Gal-HSA**/HSA	
		avg		*avg*
L1	0.528		0.256	
L2	0.641		7.331	
L3	0.802	0.716	0.686	0.348
L4	0.63		0.01	
L5	0.77	0.405	0.512	0.519
L6	0.27		0.737	
L7	0.175		0.307	

great	0 - 0.2
good	0.2 - 0.5
fair	0.5 - 0.8
poor	0.8 - >1

Figure S5. Peptide sequence of albumin residues 1-100, along with matching peptide fragments L1-L7 detected via LC-MS/MS. The inserted table shows the calculated abundance ratios, which are color-coded to assess the degree of the corresponding lysine residues as a potential ligation site.

Figure S6 shows albumin residues 121-220, along with the corresponding peptide fragments L8-L12 that have notable abundance ratio changes throughout the LC-MS/MS studies. Lysine residues of interest were identified at positions 136, 137, 159, 162, 190, 195, 199, and 205.

- For Lys136/137, peptide fragments L8 and L9 designates either one of the lysines as a fair site of ligation.
- For Lys159/162 peptide fragments L10 and L11 designates either one of the lysines as a good site of ligation.
- For Lys190/195/199/205, peptide fragment L12 designates either one of the lysines as a great site of ligation.

```
121 DVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQ
    --------DNEETFLKKYLY------------------------------ L8
    -----------ETFLKKYLY------------------------------ L9
    --------------------EIARRHPYFYAPELLFFAKRYKAAFTECCQ L10
    --------------------EIARRHPYFYAPELLFFAKRYKAAFT---- L11
    ------------------------------------------------- L12

171 AADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAFKAWAVARLS
    ------------------------------------------------- L8
    ------------------------------------------------- L9
    AA----------------------------------------------- L10
    ------------------------------------------------- L11
    ---------------DEGKASSAKQRLKCASLQKFG------------- L12
```

	Abundance Ratios			
	5Gal-HSA/HSA		**10Gal-HSA**/HSA	
		avg		*avg*
L8	0.813	0.737	0.503	0.653
L9	0.661		0.802	
L10	0.65	0.494	0.583	0.363
L11	0.337		0.142	
L12	0.175		0.171	

great	0 - 0.2
good	0.2 - 0.5
fair	0.5 - 0.8
poor	0.8 - >1

Figure S6. Peptide sequence of albumin residues 121-220, along with matching peptide fragments L8-L12 detected via LC-MS/MS. The inserted table shows the calculated abundance ratios, which are color-coded to assess the degree of the corresponding lysine residues as a potential ligation site.

Figure S7 shows albumin residues 351-450, along with the corresponding peptide fragments L13-L25 that have notable abundance ratio changes throughout the LC-MS/MS studies. Lysine residues of interest were identified at positions 389, 402, 413, 414, 432, 436, 439, and 444.

- For Lys389, peptide fragments L13-L15 designates it as a good/fair site of ligation.
- For Lys402/413/414, peptide fragments L16 and L17 designates either one of the lysines as a good site of ligation.
- For Lys432/436/439, peptide fragment L23 designates either one of the lysines as a good/great site of ligation.
- For Lys444, peptide fragments L24 and L25 designates it as a great site of ligation.

```
351 KTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGE
    ------------------------------EEPQNLIKQNCELF----- L13
    ------------------------------EEPQNLIKQNC-------- L14
    -------------------------------EPQNLIKQNC-------- L15
    ------------------------------------------------E L16
    -----------------------------------------------EQLGE L17
    -----------------------------------------------EQLGE L18
    ------------------------------------------------- L19
    ------------------------------------------------- L20
    ------------------------------------------------- L21
    ------------------------------------------------- L22
    ------------------------------------------------- L23
    ------------------------------------------------- L24
    ------------------------------------------------- L25

401 YKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAE
    ------------------------------------------------- L13
    ------------------------------------------------- L14
    ------------------------------------------------- L15
    YKFQNALLVRYTKKVPQVSTPTLV------------------------- L16
    YKFQNALLVRYTKKVPQVSTPTLV------------------------- L17
    YKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAE L18
    ------------------------EVSRNLGKVGSKCCKHPEAKRMPCAE L19
    ------------------------EVSRNLGKVGSKCCKHPEAKRMPCAE L20
    ------------------------EVSRNLGKVGSKCCKHPEAKRMPCA- L21
    ------------------------EVSRNLGKVGSKCCKHPEAKRMPCA- L22
    ------------------------EVSRNLGKVGSKCCKHP--------- L23
    ------------------------------------------EAKRMPCA- L24
    ------------------------------------------EAKRMPCAE L25
```

Figure S7. Peptide sequence of albumin residues 351-450, along with matching peptide fragments L13-25 detected via LC-MS/MS. The inserted table shows the calculated abundance ratios, which are color-coded to assess the degree of the corresponding lysine residues as a potential ligation site.

	Abundance Ratios			
	5Gal-HSA/HSA		**10Gal-HSA**/HSA	
		avg		*avg*
L13	0.565	0.482	0.653	0.664
L14	0.387		0.62	
L15	0.494		0.72	
L16	0.234	0.408	0.215	0.258
L17	0.581		0.301	
L18	0.01		0.01	
L19	0.777	0.498	0.262	0.378
L20	0.449		0.408	
L21	0.343		0.202	
L22	0.423		0.64	
L23	0.498		0.176	
L24	0.158	0.16	0.01	0.072
L25	0.162		0.134	

great	0 - 0.2
good	0.2 - 0.5
fair	0.5 - 0.8
poor	0.8 - >1

Figure S7 (cont). Peptide sequence of albumin residues 351-450, along with matching peptide fragments L13-L25 detected via LC-MS/MS. The inserted table shows the calculated abundance ratios, which are color-coded to assess the degree of the corresponding lysine residues as a potential ligation site.

Figure S8 shows albumin residues 501-550, along with the corresponding peptide fragments L26-L39 that have notable abundance ratio changes throughout the LC-MS/MS studies. Lysine residues of interest were identified at positions 519, 524, 525, 534, 536, 538, and 541.

- Because peptide fragments L26-L35 has overlaps with too many potential lysine sites of conjugation, their analysis is strictly limited to looking at them in bulk. The general low abundance ratios found with these peptide fragments suggest that this portion of the albumin protein is strongly favorable for conjugation with the glycan–aldehyde probes.
- For Lys519, peptide fragments L38 and L39 designates it as a good site of ligation.
- For Lys524/525, peptide fragment L37 designates either one of the lysines as a good/great site of ligation.
- For Lys534/536/538/541, peptide fragment L36 designates either one of the lysines as a great site of ligation.

```
501 EFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDD
    -----------DICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVM-- L26
    -----------DICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVM-- L27
    -----------------EKERQIKKQTALVELVKHKPKATKEQLKAVM-- L28
    -----------------EKERQIKKQTALVELVKHKPKATKEQLKAVM-- L29
    -------------------ERQIKKQTALVELVKHKPKATKEQLKAVM-- L30
    -----------------EKERQIKKQTALVELVKHKPKATK--------- L31
    -------------------ERQIKKQTALVELVKHKPKATK--------- L32
    -----------DICTLSEKERQIKKQTALVELVKHKPKATK--------- L33
    -----------------EKERQIKKQTALVELV----------------- L34
    -----------DICTLSEKERQIKKQTALV-------------------- L35
    ------------------------------ELVKHKPKATK--------- L36
    -------------------ERQIKKQTALVELV----------------- L37
    EFNAETFTFHADICTLSEK------------------------------- L38
    -----------DICTLSEK------------------------------- L39
```

	Abundance Ratios			
	5Gal-HSA/HSA		**10Gal-HSA**/HSA	
		avg		*avg*
L26	0.114	0.162	0.01	0.045
L27	0.514		0.213	
L28	0.21		0.01	
L29	0.017		0.01	
L30	0.01		0.01	
L31	0.078		0.02	
L32	0.103		0.01	
L33	0.124		0.022	
L34	0.251		0.039	
L35	0.2		0.101	
L36	0.197		0.129	
L37	0.214		0.188	
L38	0.067	0.433	0.104	0.486
L39	0.799		0.868	

great	0 - 0.2
good	0.2 - 0.5
fair	0.5 - 0.8
poor	0.8 - >1

Figure S8. Peptide sequence of albumin residues 501-550, along with matching peptide fragments L26-L39 detected via LC-MS/MS. The inserted table shows the calculated abundance ratios, which are color-coded to assess the degree of the corresponding lysine residues as a potential ligation site.

DKLCTVATLRETYGEMA (72-88)

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

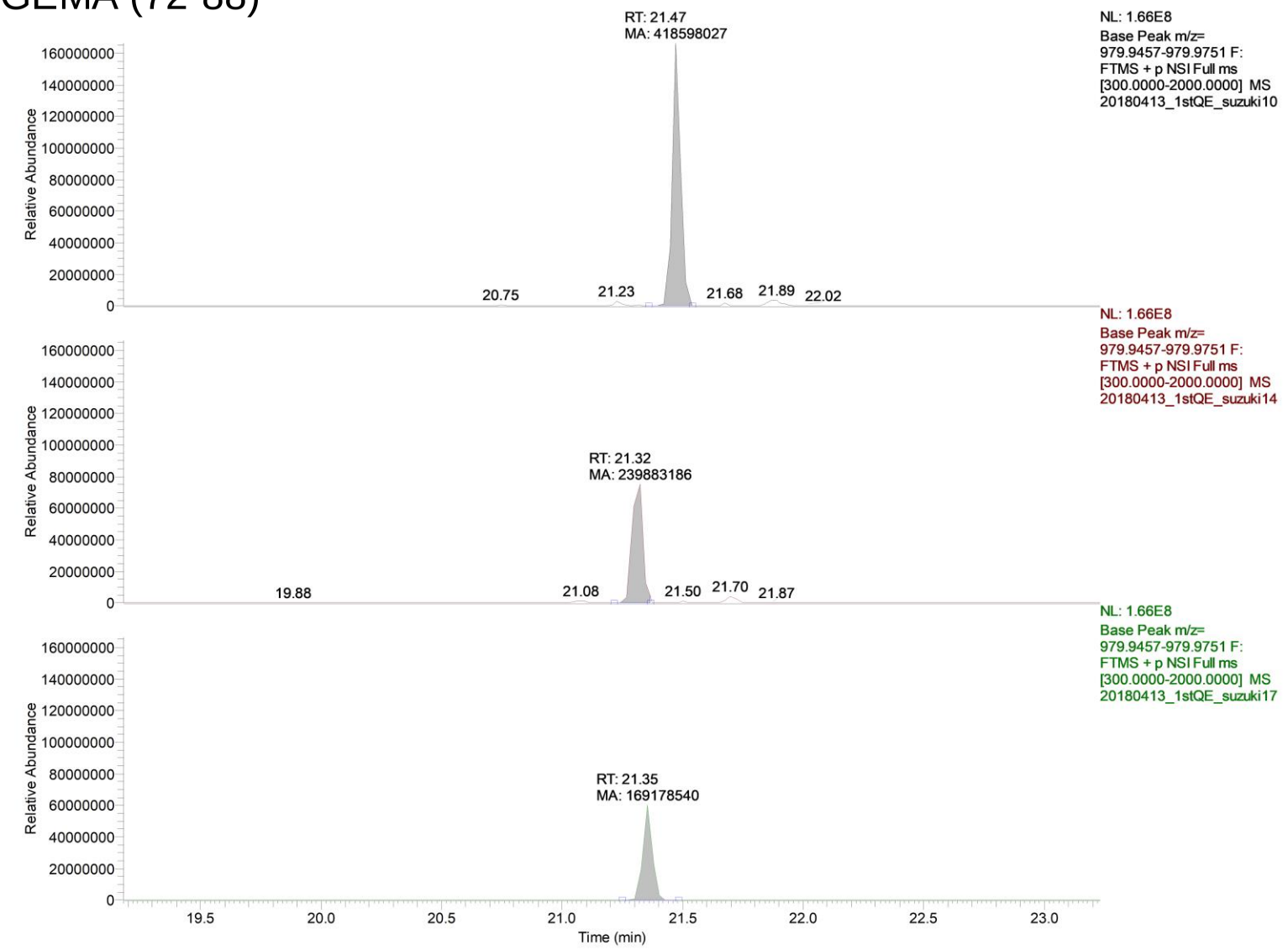


Figure S9. Magnified MS chromatogram (retention time: 19.18 to 23.23 min) on the peptide fragment [DKLCTVATLRETYGEMA] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

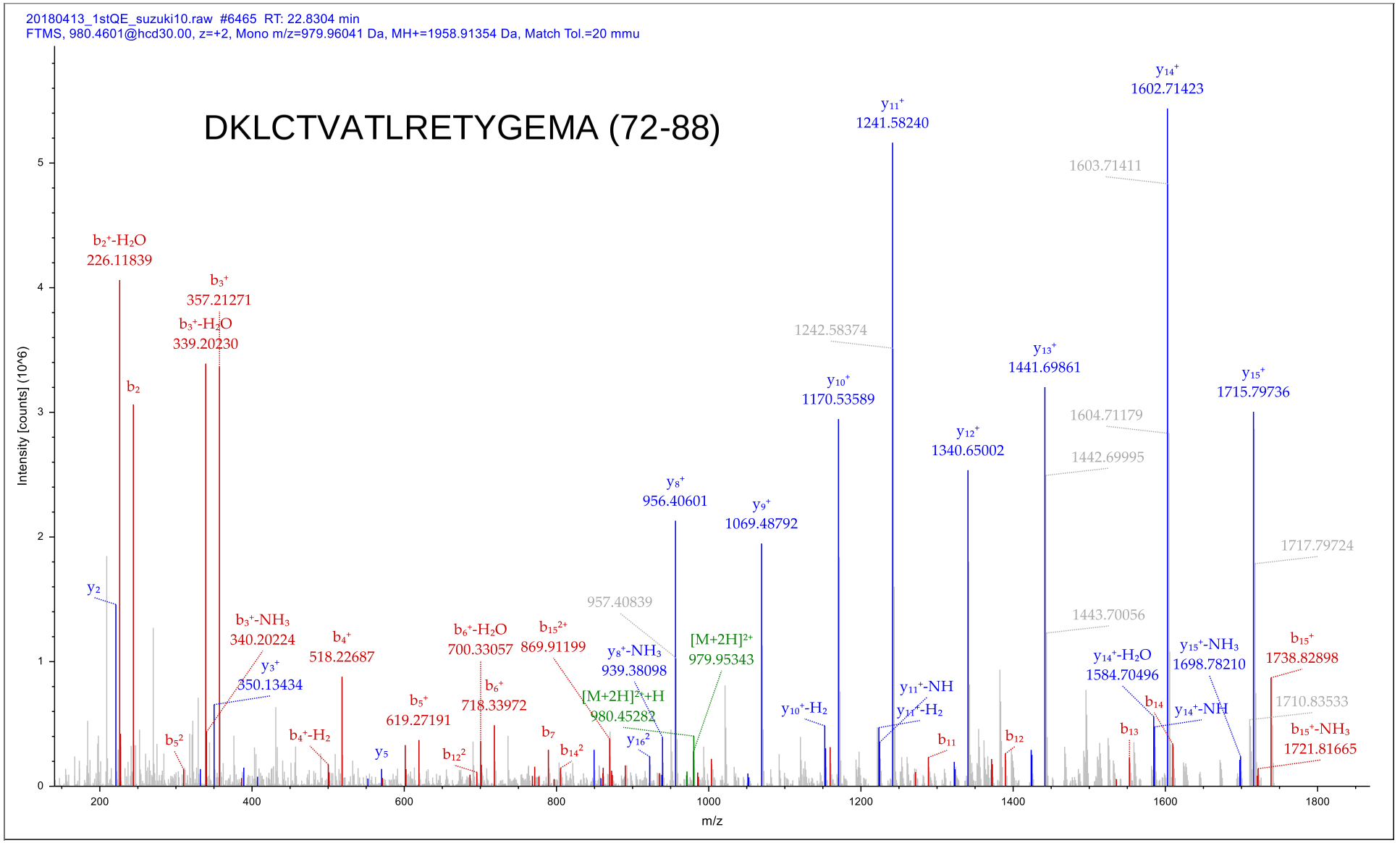


Figure S10. MS/MS spectra of the peptide fragment [DKLCTVATLRETYGEMA] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

DKLCTVATLRETYGEMA (72-88)

#1	b^{+}	b^{2+}	Seq.	y^{+}	y^{2+}	#2
1	116.0342	58.5208	D			17
2	244.1292	122.5682	K	1843.8929	922.4501	16
3	357.2133	179.1103	L	1715.7979	858.4026	15
4	518.2279	259.6176	C-Carboxymethyl	1602.7138	801.8606	14
5	619.2756	310.1414	T	1441.6992	721.3532	13
6	718.3440	359.6756	V	1340.6515	670.8294	12
7	789.3811	395.1942	A	1241.5831	621.2952	11
8	890.4288	445.7180	T	1170.5460	585.7766	10
9	1003.5129	502.2601	L	1069.4983	535.2528	9
10	1159.6140	580.3106	R	956.4142	478.7107	8
11	1288.6566	644.8319	E	800.3131	400.6602	7
12	1389.7042	695.3558	T	671.2705	336.1389	6
13	1552.7676	776.8874	Y	570.2228	285.6151	5
14	1609.7890	805.3982	G	407.1595	204.0834	4
15	1738.8316	869.9195	E	350.1380	175.5727	3
16	1869.8721	935.4397	M	221.0954	111.0514	2
17			A	90.0550	45.5311	1

Figure S11. Summary of detected fragment matches for the peptide [DKLCTVATLRETYGEMA].

DEGKASSAKQRLKCASLQKFG (187-207)

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

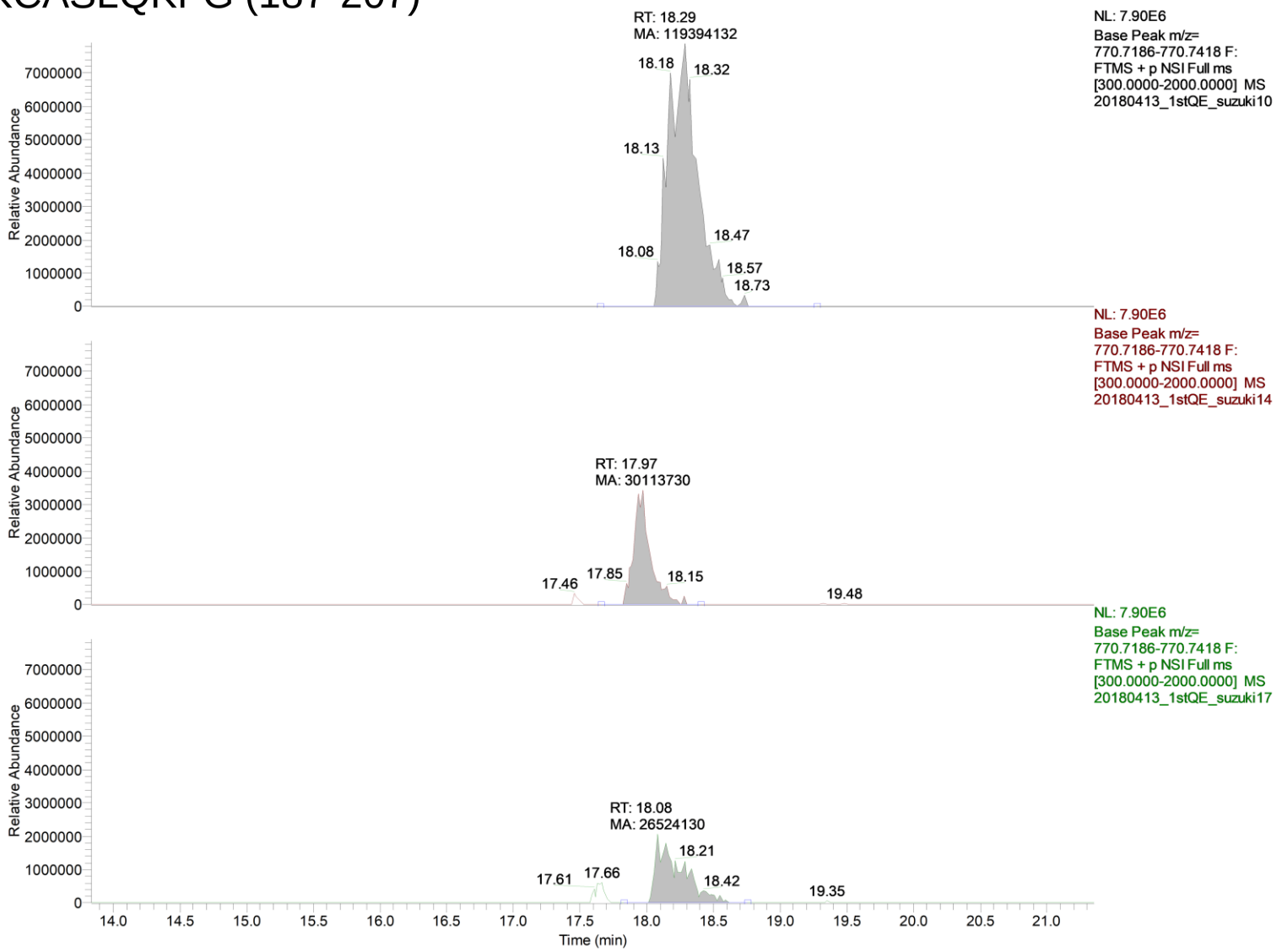


Figure S12. Magnified MS chromatogram (retention time: 13.83 to 21.35 min) on the peptide fragment [DEGKASSAKQRLKCASLQKFG] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

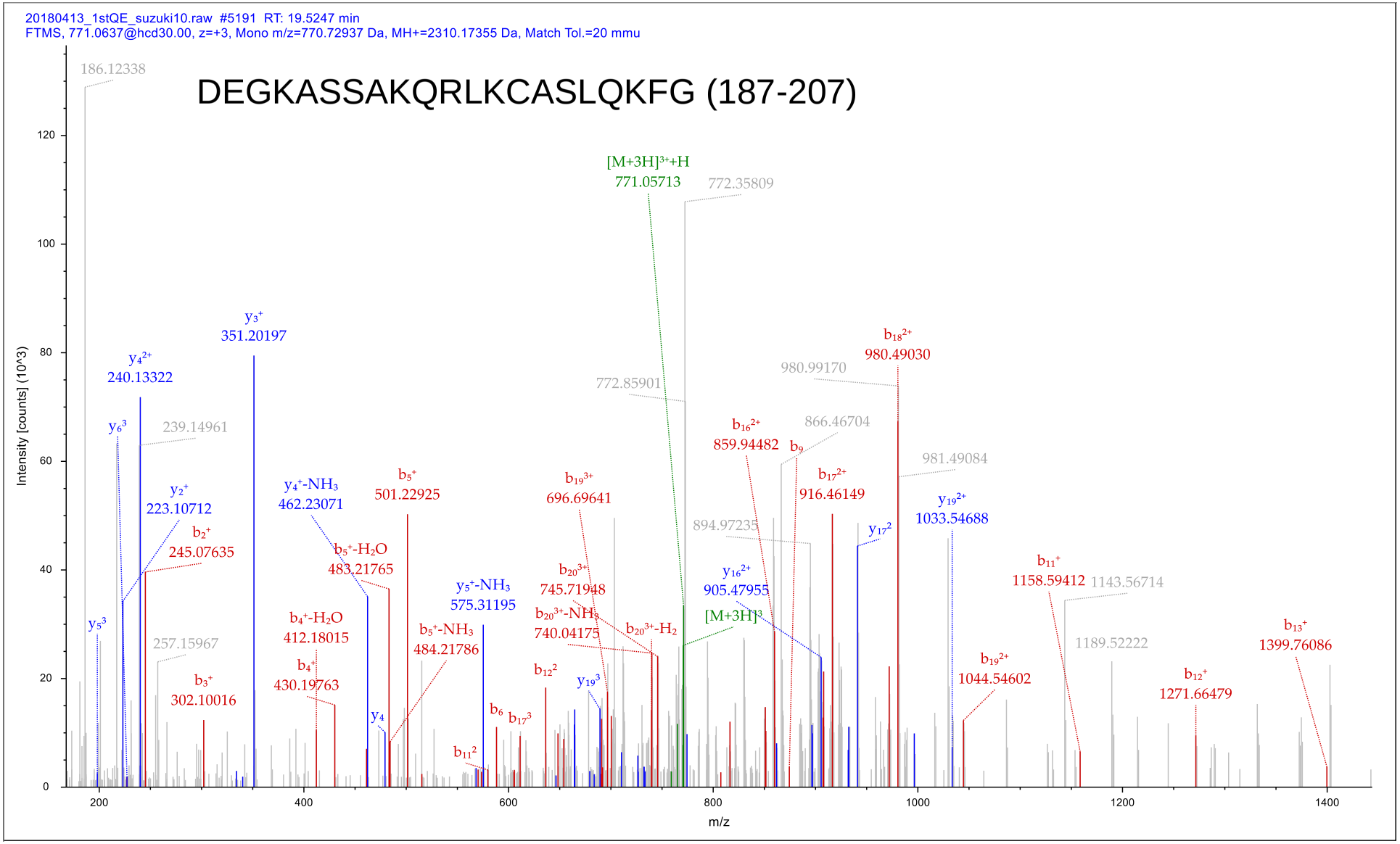


Figure S13. MS/MS spectra of the peptide fragment [DEGKASSAKQRLKCASLQKFG] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

DEGKASSAKQRLKCASLQKFG (187-207)

#1	b^{+}	b^{2+}	b^{3+}	Seq.	y^{+}	y^{2+}	y^{3+}	#2
1	116.0342	58.5208	39.3496	D				21
2	245.0768	123.0420	82.3638	E	2195.1601	1098.0837	732.3916	20
3	302.0983	151.5528	101.3709	G	2066.1175	1033.5624	689.3774	19
4	430.1932	215.6003	144.0693	K	2009.0961	1005.0517	670.3702	18
5	501.2304	251.1188	167.7483	A	1881.0011	941.0042	627.6719	17
6	588.2624	294.6348	196.7590	S	1809.9640	905.4856	603.9928	16
7	675.2944	338.1508	225.7697	S	1722.9320	861.9696	574.9822	15
8	746.3315	373.6694	249.4487	A	1635.8999	818.4536	545.9715	14
9	874.4265	437.7169	292.1470	K	1564.8628	782.9350	522.2925	13
10	1002.4851	501.7462	334.8332	Q	1436.7678	718.8876	479.5941	12
11	1158.5862	579.7967	386.8669	R	1308.7093	654.8583	436.9079	11
12	1271.6702	636.3388	424.5616	L	1152.6082	576.8077	384.8742	10
13	1399.7652	700.3862	467.2599	K	1039.5241	520.2657	347.1796	9
14	1560.7799	780.8936	520.9315	C-Carboxymethyl	911.4291	456.2182	304.4812	8
15	1631.8170	816.4121	544.6105	A	750.4145	375.7109	250.8097	7
16	1718.8490	859.9281	573.6212	S	679.3774	340.1923	227.1306	6
17	1831.9331	916.4702	611.3159	L	592.3453	296.6763	198.1200	5
18	1959.9917	980.4995	654.0021	Q	479.2613	240.1343	160.4253	4
19	2088.0866	1044.5469	696.7004	K	351.2027	176.1050	117.7391	3
20	2235.1550	1118.0812	745.7232	F	223.1077	112.0575	75.0408	2
21				G	76.0393	38.5233	26.0180	1

Figure S14. Summary of detected fragment matches for the peptide [DEGKASSAKQRLKCASLQKFG].

EVSRNLGKVGSKCCKHPEAKRMPCA (425-449)

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

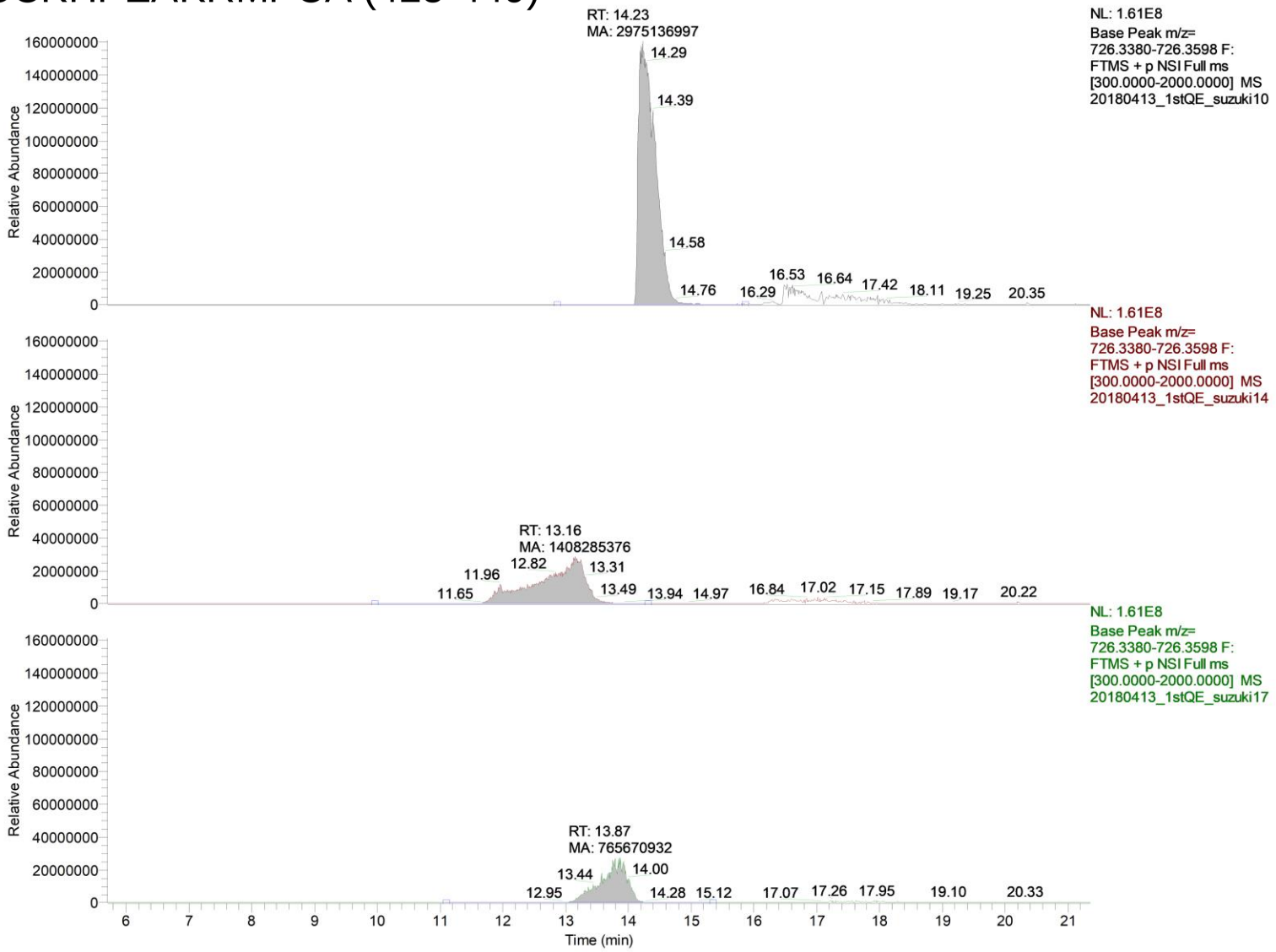


Figure S15. Magnified MS chromatogram (retention time: 5.70 to 21.35 min) on the peptide fragment [EVSRNLGKVGSKCCKHPEAKRMPCA] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

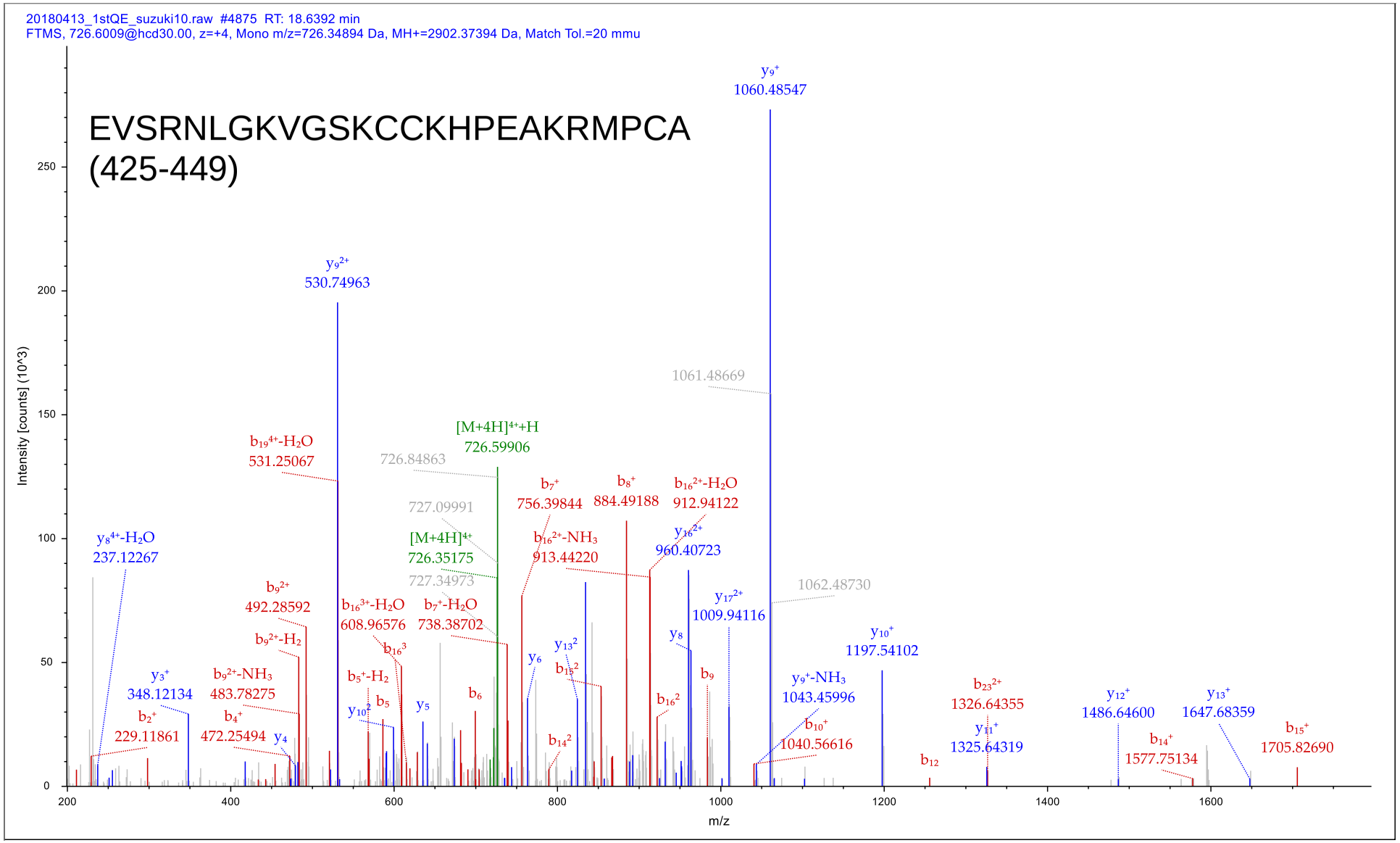


Figure S16. MS/MS spectra of the peptide fragment [EVSRNLGKVGSKCCKHPEAKRMPCA] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

EVSRNLGKVGSKCCKHPEAKRMPCA (425-449)

#1	b^{+}	b^{2+}	b^{3+}	b^{4+}	Seq.	y^{+}	y^{2+}	y^{3+}	y^{4+}	#2
1	130.0499	65.5286	44.0215	33.2679	E					25
2	229.1183	115.0628	77.0443	58.0350	V	2773.3365	1387.1719	925.1170	694.0896	24
3	316.1503	158.5788	106.0550	79.7930	S	2674.2680	1337.6377	892.0942	669.3225	23
4	472.2514	236.6294	158.0887	118.8183	R	2587.2360	1294.1216	863.0835	647.5645	22
5	586.2944	293.6508	196.1030	147.3291	N	2431.1349	1216.0711	811.0498	608.5392	21
6	699.3784	350.1929	233.7977	175.6001	L	2317.0920	1159.0496	773.0355	580.0285	20
7	756.3999	378.7036	252.8048	189.8554	G	2204.0079	1102.5076	735.3408	551.7574	19
8	884.4948	442.7511	295.5031	221.8792	K	2146.9864	1073.9969	716.3337	537.5021	18
9	983.5633	492.2853	328.5259	246.6463	V	2018.8915	1009.9494	673.6354	505.4783	17
10	1040.5847	520.7960	347.5331	260.9016	G	1919.8231	960.4152	640.6125	480.7112	16
11	1127.6168	564.3120	376.5438	282.6596	S	1862.8016	931.9044	621.6054	466.4559	15
12	1255.7117	628.3595	419.2421	314.6834	K	1775.7696	888.3884	592.5947	444.6979	14
13	1416.7264	708.8668	472.9136	354.9371	C-Carboxymethyl	1647.6746	824.3409	549.8964	412.6741	13
14	1577.7410	789.3742	526.5852	395.1907	C-Carboxymethyl	1486.6600	743.8336	496.2248	372.4204	12
15	1705.8360	853.4216	569.2835	427.2145	K	1325.6453	663.3263	442.5533	332.1668	11
16	1842.8949	921.9511	614.9698	461.4792	H	1197.5503	599.2788	399.8550	300.1430	10
17	1939.9477	970.4775	647.3207	485.7424	P	1060.4914	530.7493	354.1687	265.8783	9
18	2068.9903	1034.9988	690.3349	518.0030	E	963.4387	482.2230	321.8177	241.6151	8
19	2140.0274	1070.5173	714.0140	535.7623	A	834.3961	417.7017	278.8035	209.3545	7
20	2268.1223	1134.5648	756.7123	567.7860	K	763.3589	382.1831	255.1245	191.5952	6
21	2424.2235	1212.6154	808.7460	606.8113	R	635.2640	318.1356	212.4262	159.5715	5
22	2555.2639	1278.1356	852.4262	639.5714	M	479.1629	240.0851	160.3925	120.5462	4
23	2652.3167	1326.6620	884.7771	663.8346	P	348.1224	174.5648	116.7123	87.7861	3
24	2813.3314	1407.1693	938.4486	704.0883	C-Carboxymethyl	251.0696	126.0385	84.3614	63.5229	2
25					A	90.0550	45.5311	30.6898	23.2692	1

Figure S17. Summary of detected fragment matches for the peptide [EVSRNLGKVGSKCCKHPEAKRMPCA].

EKERQIKKQTALVELVKHKPKATK (518-541)

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

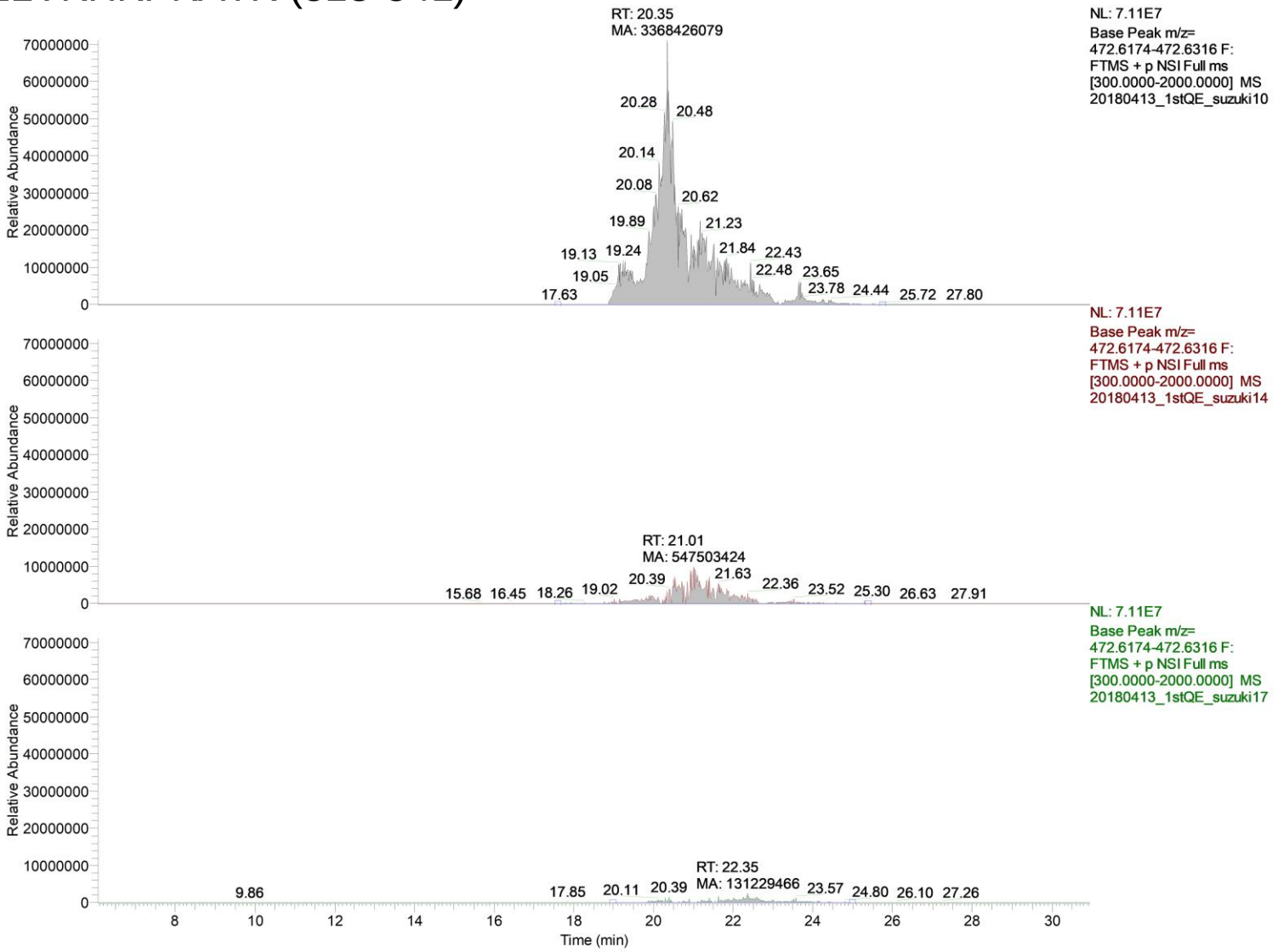


Figure S18. Magnified MS chromatogram (retention time: 6.05 to 30.94 min) on the peptide fragment [EKERQIKKQTALVELVKHKPKATK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

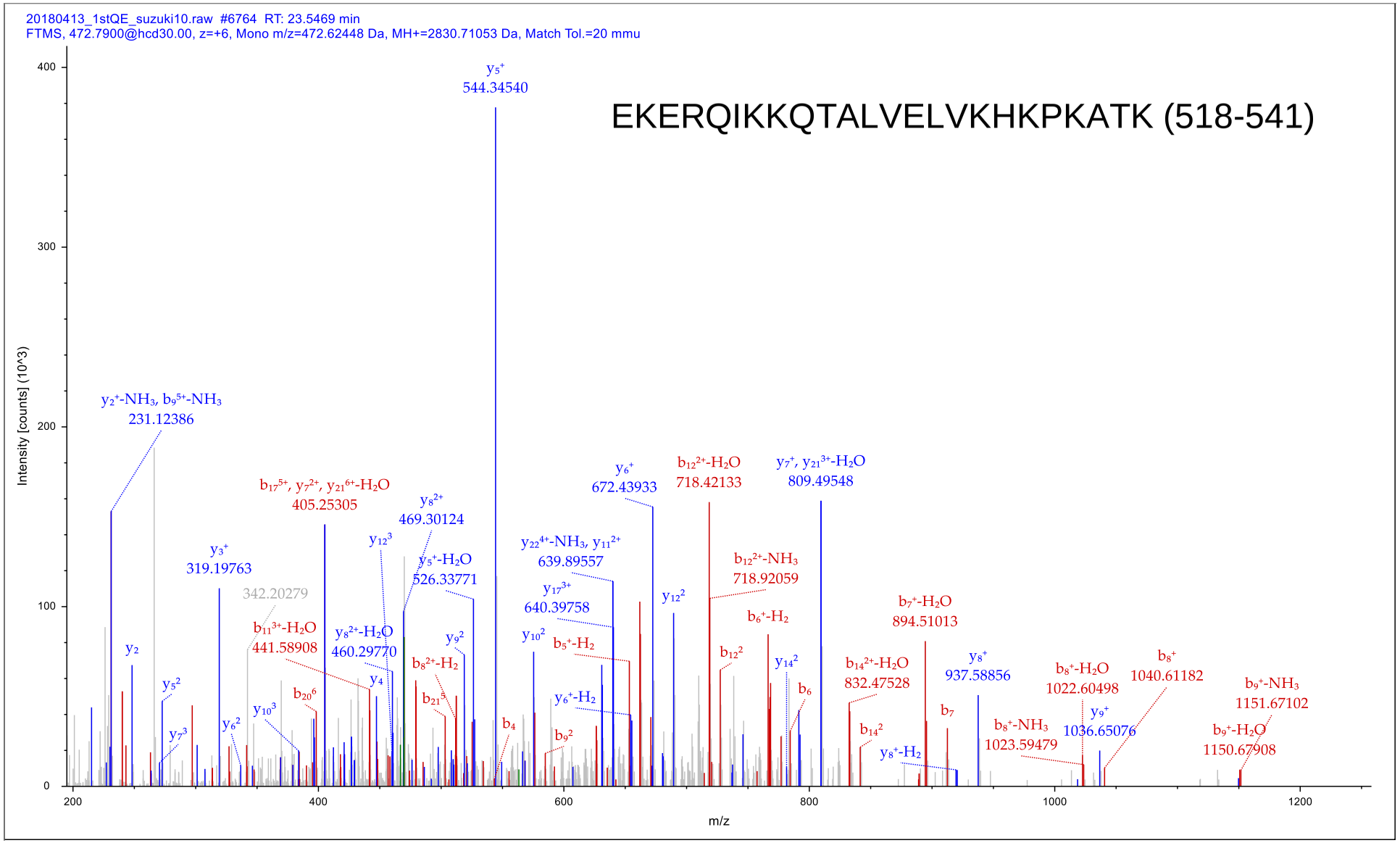


Figure S19. MS/MS spectra of the peptide fragment [EKERQIKKQTALVELVKHKPKATK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

EKERQIKKQTALVELVKHKPKATK (518-541)

#1	b^{+}	b^{2+}	b^{3+}	b^{4+}	b^{5+}	b^{6+}	Seq.	y^{+}	y^{2+}	y^{3+}	y^{4+}	y^{5+}	y^{6+}	#2
1	130.0499	65.5286	44.0215	33.2679	26.8158	22.5144	E							24
2	258.1448	129.5761	86.7198	65.2917	52.4348	43.8635	K	2701.6563	1351.3318	901.2236	676.1695	541.1371	451.1155	23
3	387.1874	194.0974	129.7340	97.5523	78.2433	65.3706	E	2573.5614	1287.2843	858.5253	644.1458	515.5181	429.7663	22
4	543.2885	272.1479	181.7677	136.5776	109.4635	91.3875	R	2444.5188	1222.7630	815.5111	611.8852	489.7096	408.2592	21
5	671.3471	336.1772	224.4539	168.5922	135.0752	112.7306	Q	2288.4177	1144.7125	763.4774	572.8599	458.4894	382.2423	20
6	784.4312	392.7192	262.1486	196.8633	157.6921	131.5779	I	2160.3591	1080.6832	720.7912	540.8452	432.8776	360.8992	19
7	912.5261	456.7667	304.8469	228.8870	183.3111	152.9271	K	2047.2750	1024.1411	683.0965	512.5742	410.2608	342.0519	18
8	1040.6211	520.8142	347.5452	260.9107	208.9300	174.2763	K	1919.1800	960.0937	640.3982	480.5505	384.6418	320.7027	17
9	1168.6797	584.8435	390.2314	292.9254	234.5418	195.6193	Q	1791.0851	896.0462	597.6999	448.5267	359.0228	299.3536	16
10	1269.7274	635.3673	423.9140	318.1873	254.7513	212.4606	T	1663.0265	832.0169	555.0137	416.5121	333.4111	278.0105	15
11	1340.7645	670.8859	447.5930	335.9466	268.9587	224.3001	A	1561.9788	781.4931	521.3311	391.2502	313.2016	261.1692	14
12	1453.8485	727.4279	485.2877	364.2176	291.5755	243.1475	L	1490.9417	745.9745	497.6521	373.4909	298.9942	249.3297	13
13	1552.9170	776.9621	518.3105	388.9847	311.3892	259.6589	V	1377.8577	689.4325	459.9574	345.2199	276.3774	230.4823	12
14	1681.9595	841.4834	561.3247	421.2453	337.1977	281.1660	E	1278.7892	639.8983	426.9346	320.4528	256.5637	213.9709	11
15	1795.0436	898.0254	599.0194	449.5164	359.8145	300.0133	L	1149.7466	575.3770	383.9204	288.1921	230.7552	192.4638	10
16	1894.1120	947.5597	632.0422	474.2835	379.6282	316.5247	V	1036.6626	518.8349	346.2257	259.9211	208.1383	173.6165	9
17	2022.2070	1011.6071	674.7405	506.3072	405.2472	337.8739	K	937.5942	469.3007	313.2029	235.1540	188.3247	157.1051	8
18	2159.2659	1080.1366	720.4268	540.5719	432.6590	360.7171	H	809.4992	405.2532	270.5046	203.1303	162.7057	135.7559	7
19	2287.3609	1144.1841	763.1251	572.5957	458.2780	382.0662	K	672.4403	336.7238	224.8183	168.8655	135.2939	112.9128	6
20	2384.4136	1192.7105	795.4761	596.8589	477.6886	398.2417	P	544.3453	272.6763	182.1200	136.8418	109.6749	91.5636	5
21	2512.5086	1256.7579	838.1744	628.8826	503.3075	419.5908	K	447.2926	224.1499	149.7690	112.5786	90.2643	75.3882	4
22	2583.5457	1292.2765	861.8534	646.6419	517.5150	431.4304	A	319.1976	160.1024	107.0707	80.5549	64.6453	54.0390	3
23	2684.5934	1342.8003	895.5360	671.9038	537.7245	448.2716	T	248.1605	124.5839	83.3917	62.7956	50.4379	42.1995	2
24							K	147.1128	74.0600	49.7091	37.5337	30.2284	25.3582	1

Figure S20. Summary of detected fragment matches for the peptide [EKERQIKKQTALVELVKHKPKATK].

1.4.4 MALDI-TOF/MS mass spectrometry

For the MALDI-TOF/MS study, three different protein samples were digested and analyzed (native HSA, **5Gal-HSA**, and **10Gal-HSA**). Trypsin was employed for enzymatic digestion, which acts by selectively cleaving peptide bonds on the C-terminal side of lysine (Lys, K) and arginine (Arg, R) residues.

Data analysis identified peptide fragments M1-M10 to be bound with the galactose-terminated glycan, which was found with varying intensity levels in both **5Gal-HSA** and **10Gal-HSA** samples. In addition, these fragments were absent from the native HSA-digested control spectrum. A summary of peptide fragments M1-M10 and the proposed lysine residues (shown in orange) that act as ligation sites are shown in Figure S21. Detailed MALDI-TOF spectrum for these peptides can be found in Figures S22-32.

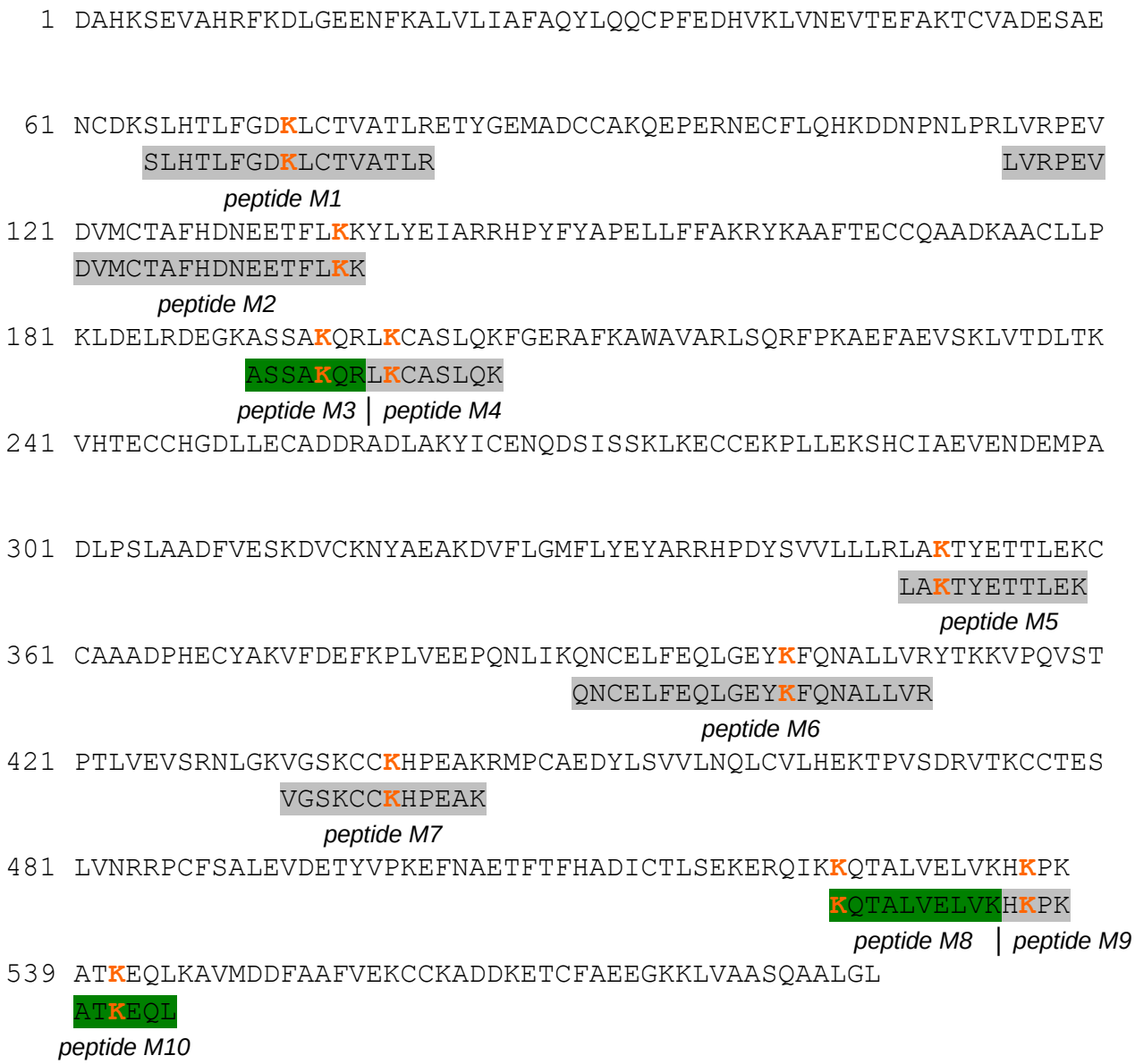


Figure S21. Summary of probable lysine residues in human serum albumin that are prone to ligation with glycan–aldehyde probes obtained via MALDI-TOF/MS methods.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

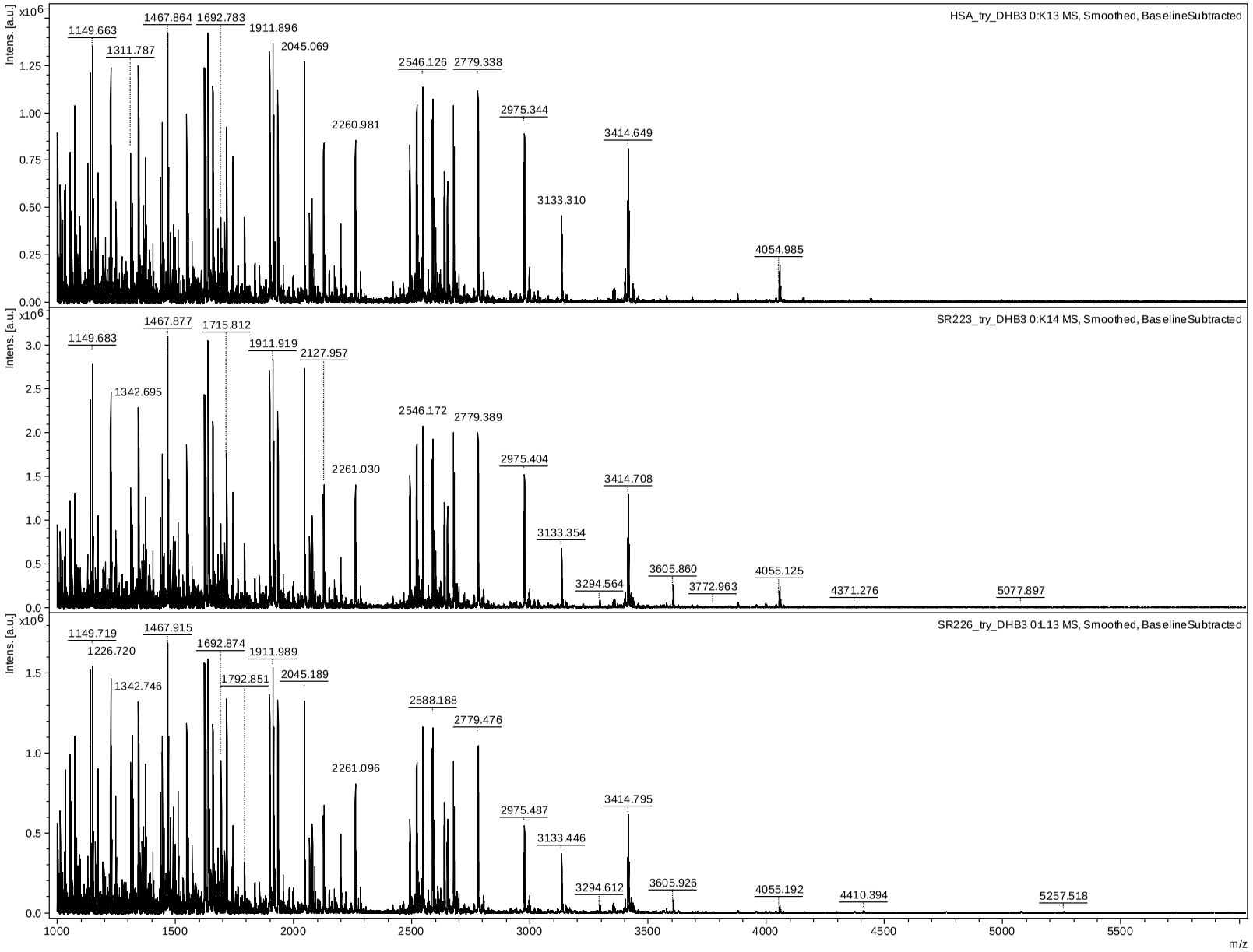


Figure S22. Full MALDI-TOF spectrum profile for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA** following trypsin digestion.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

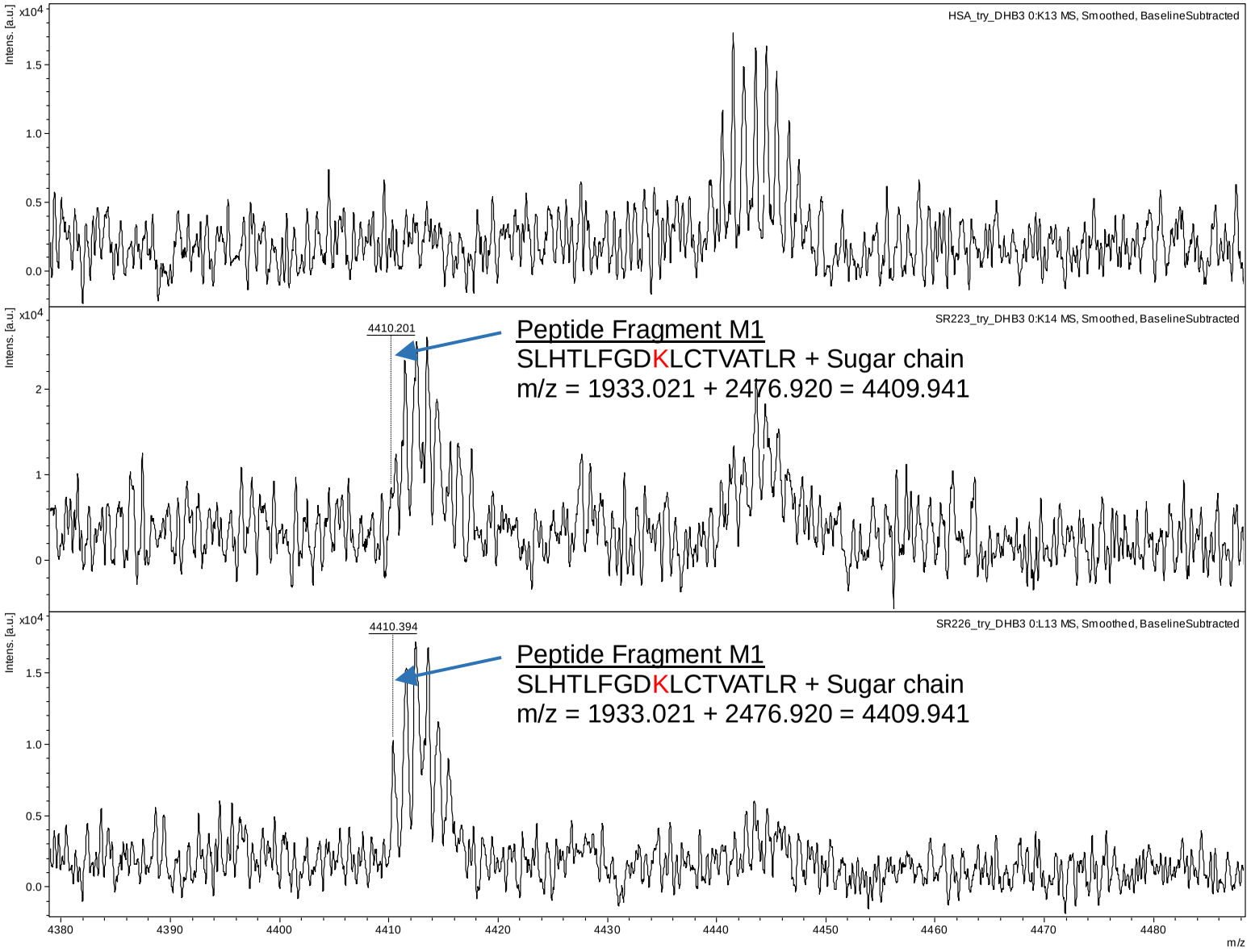


Figure S23. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M1 [SLHTLFGDKLCTVATLR] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

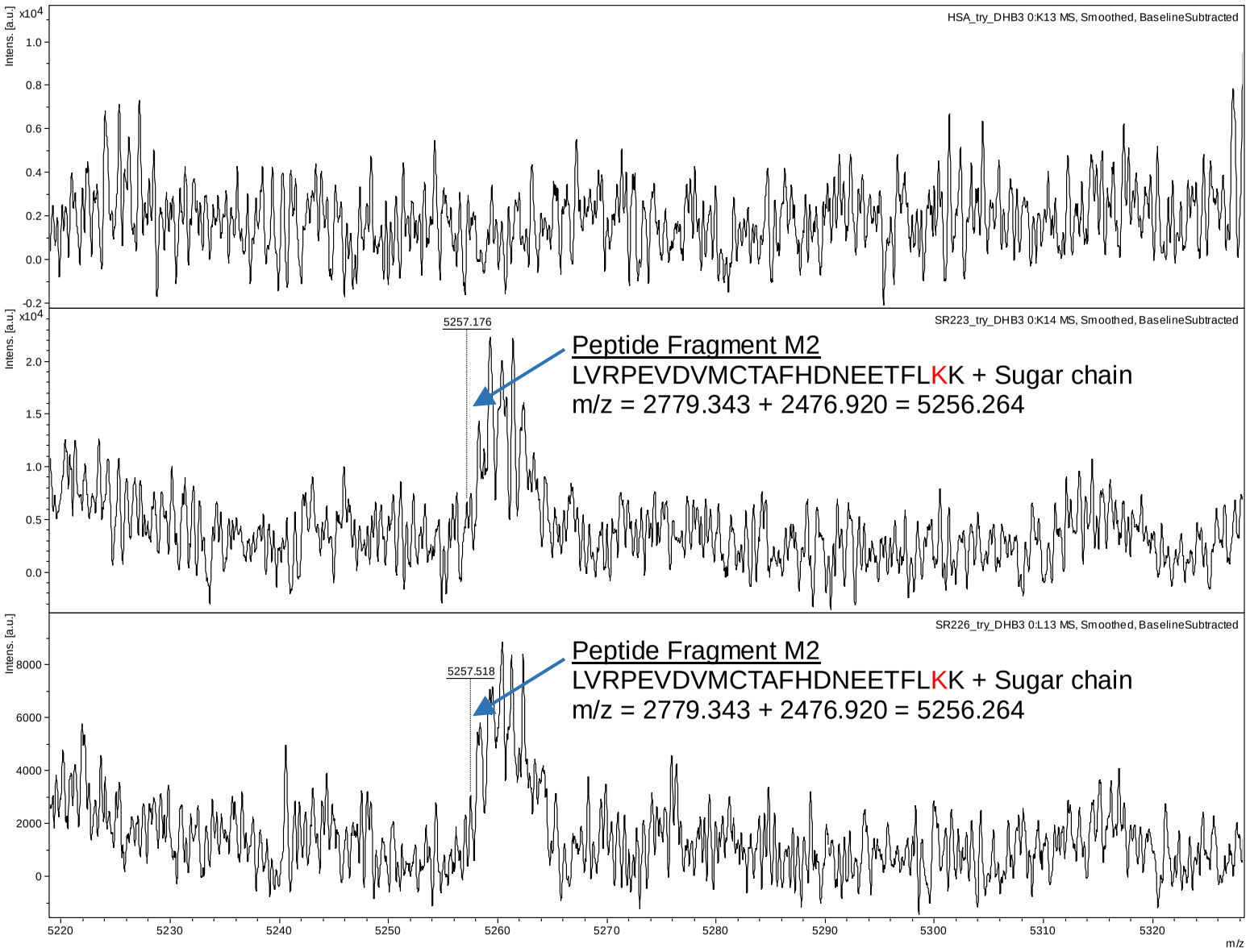


Figure S24. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M2 [LVRPEVDVMCTAFHDNEETFLKK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

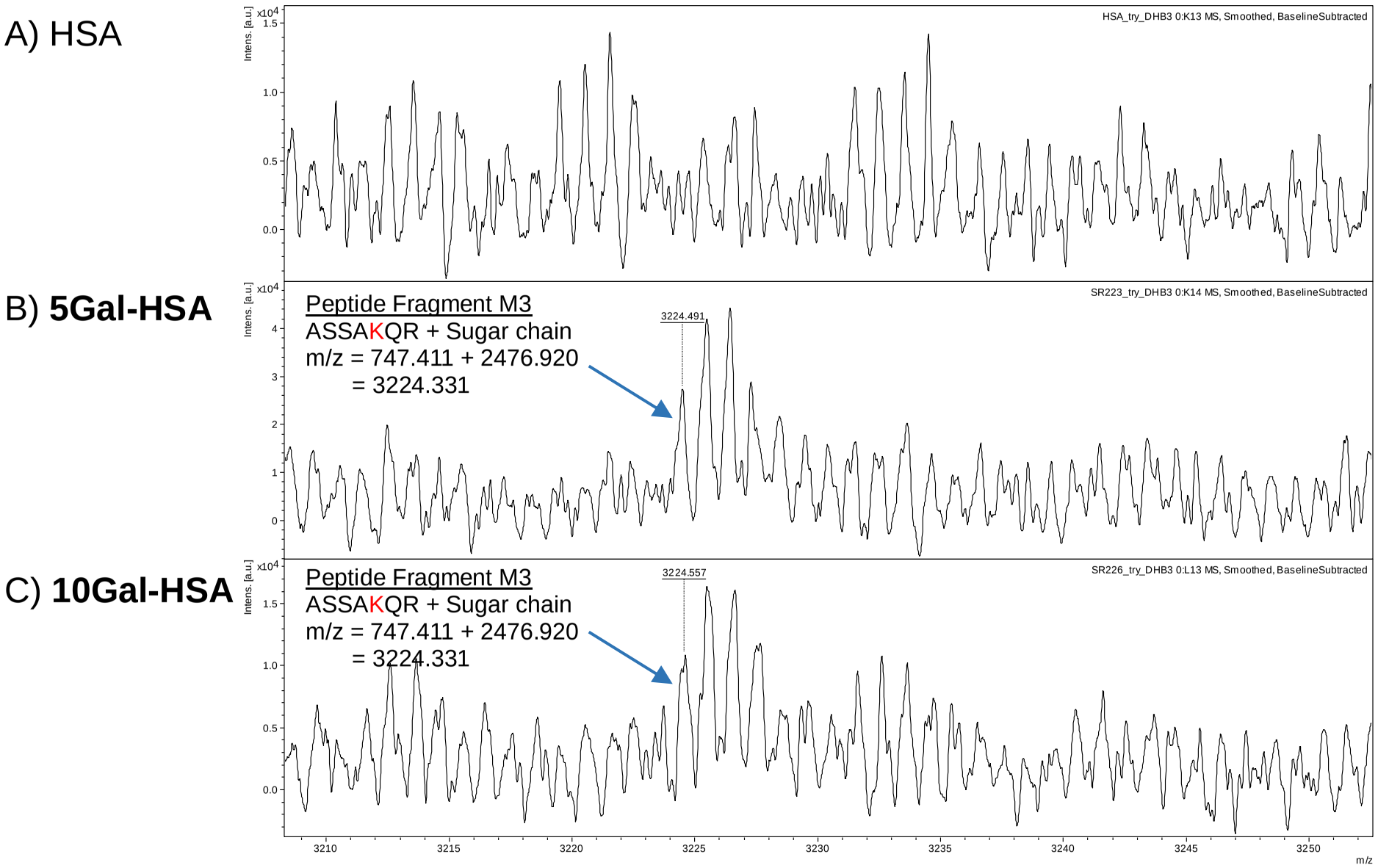


Figure S25. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M3 [ASSAKQR] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

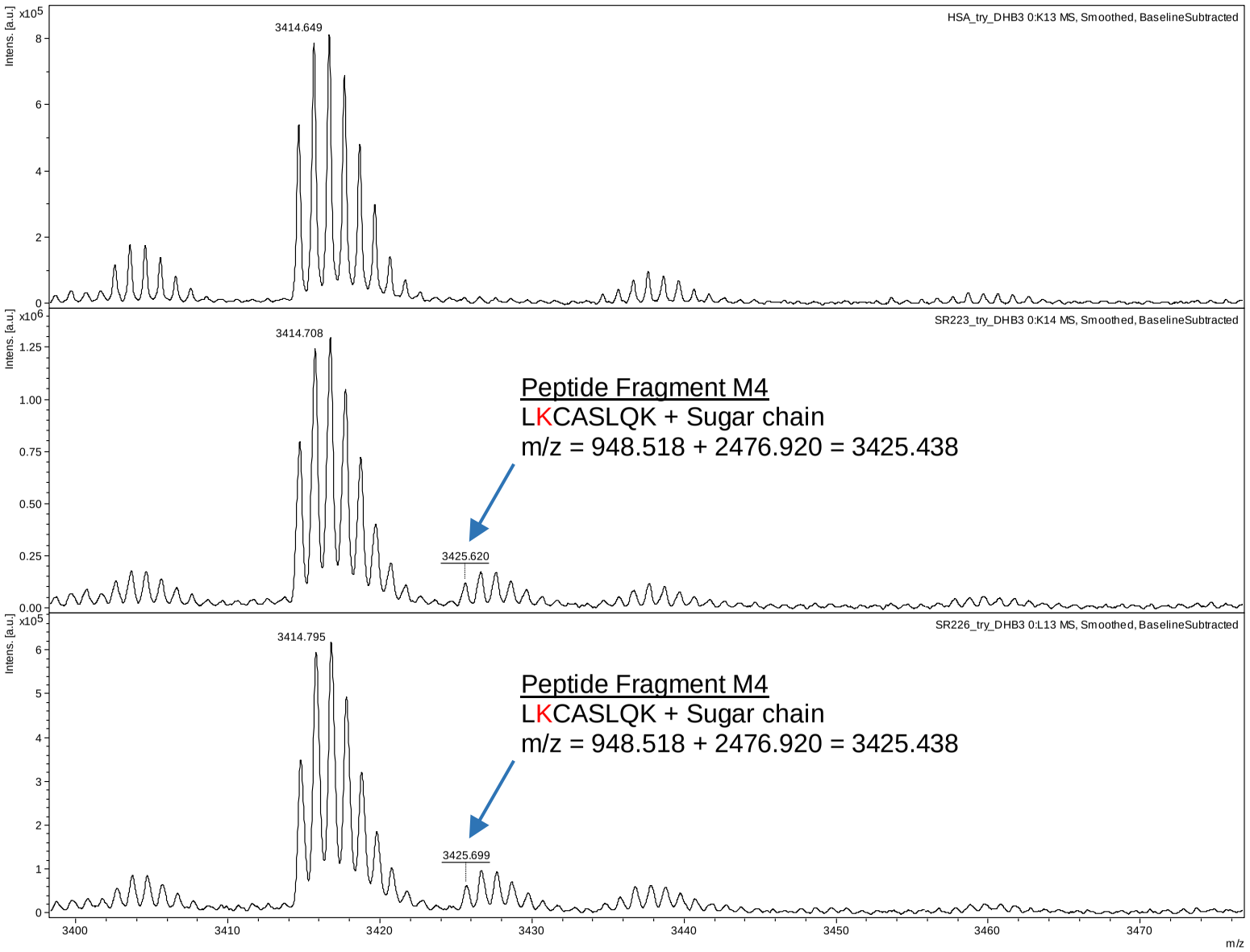


Figure S26. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M4 [LKCASLQK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

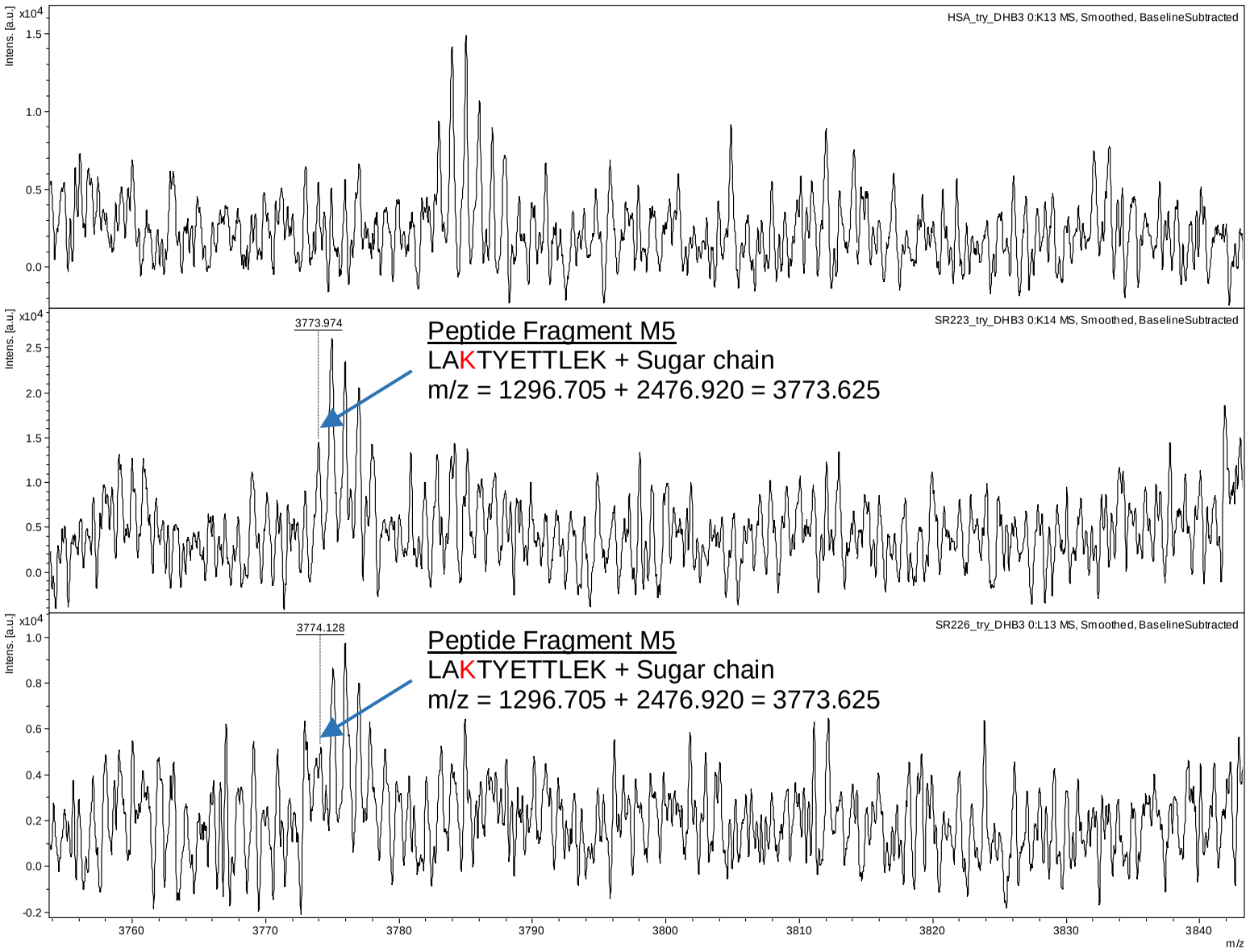


Figure S27. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M5 [LAKTYETTLEK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

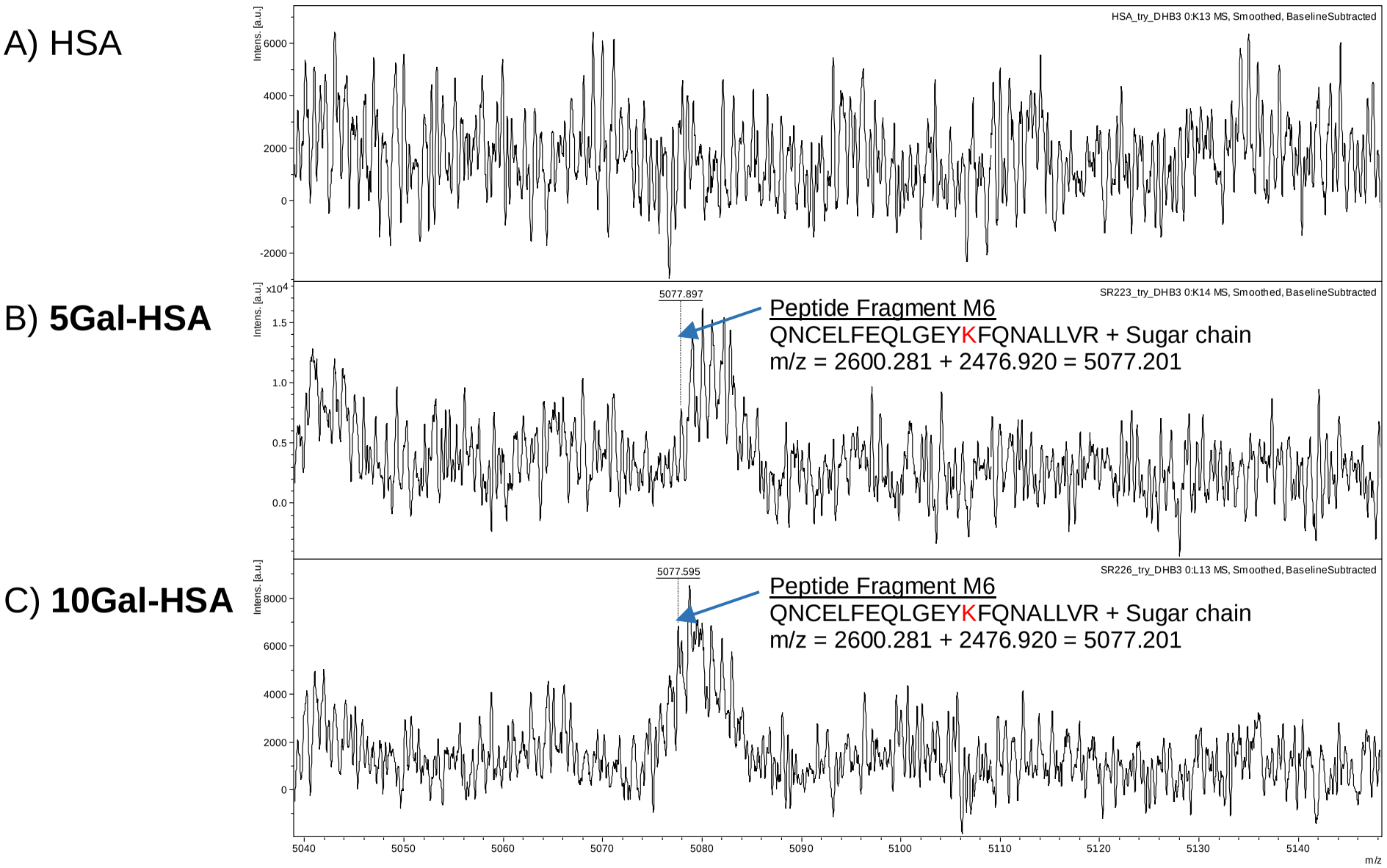


Figure S28. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M6 [QNCELFEQLGEYKFQNALLVR] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

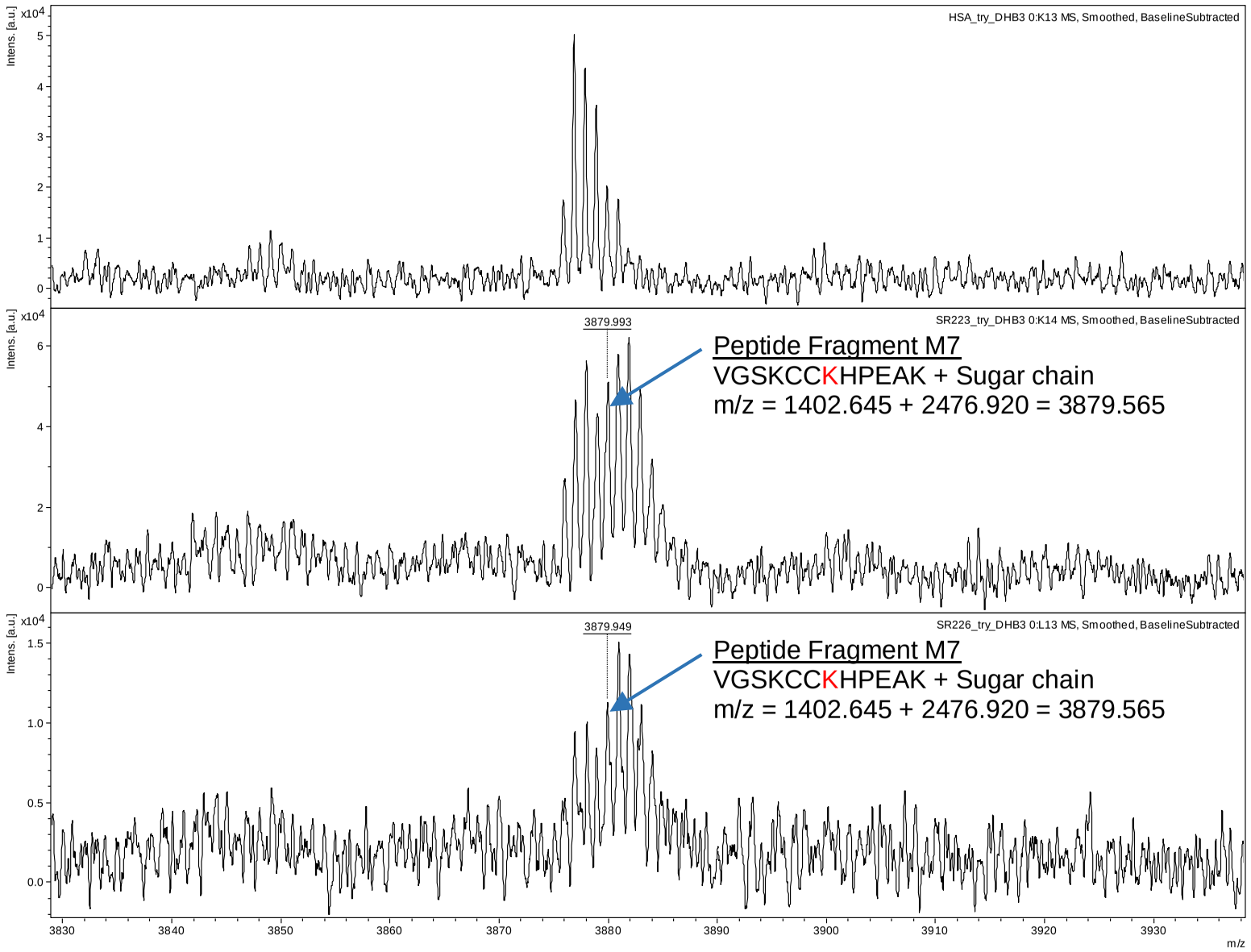


Figure S29. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M7 [VGSKCCKHPEAK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

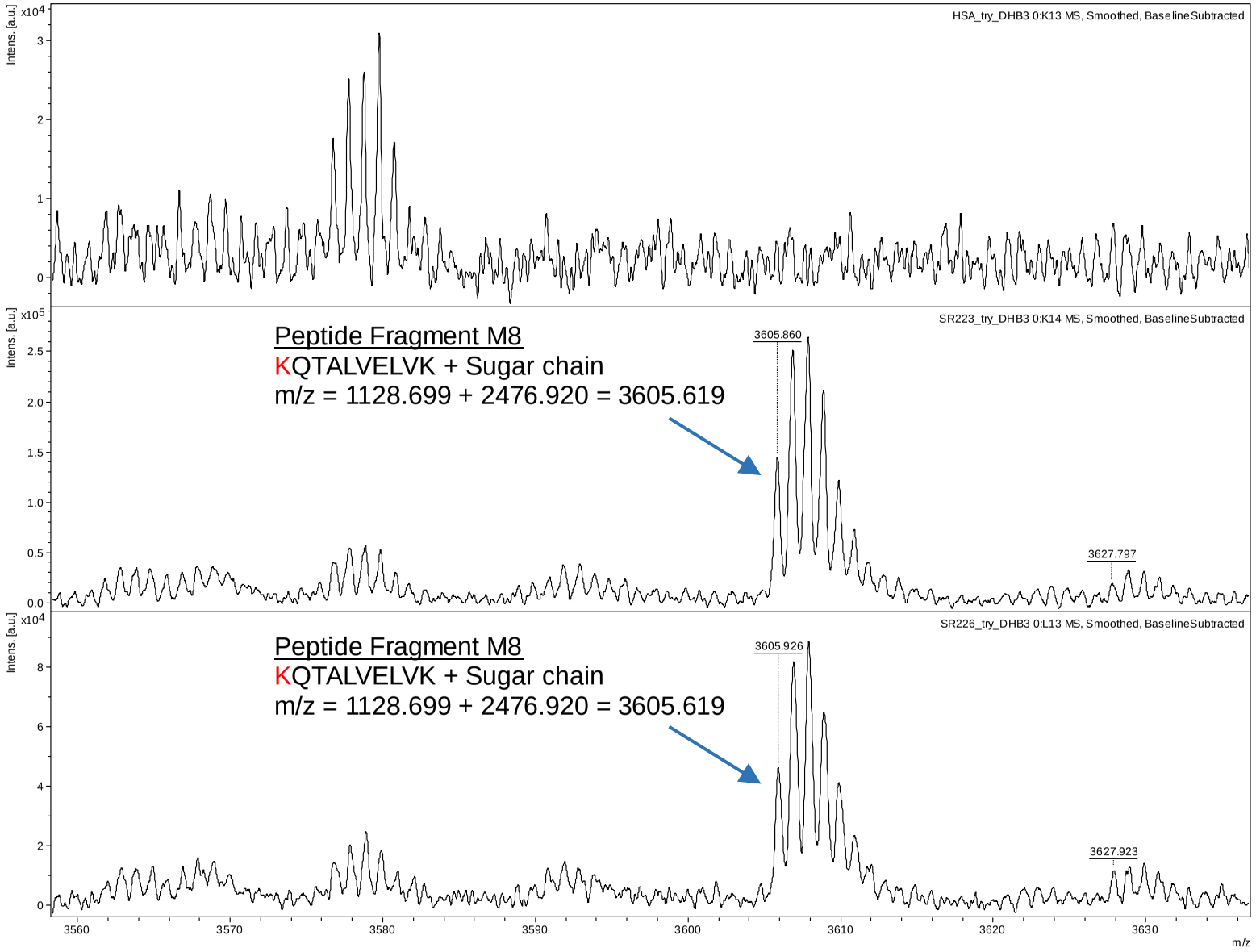


Figure S30. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M8 [KQTALVELVK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

A) HSA

B) **5Gal-HSA**

C) **10Gal-HSA**

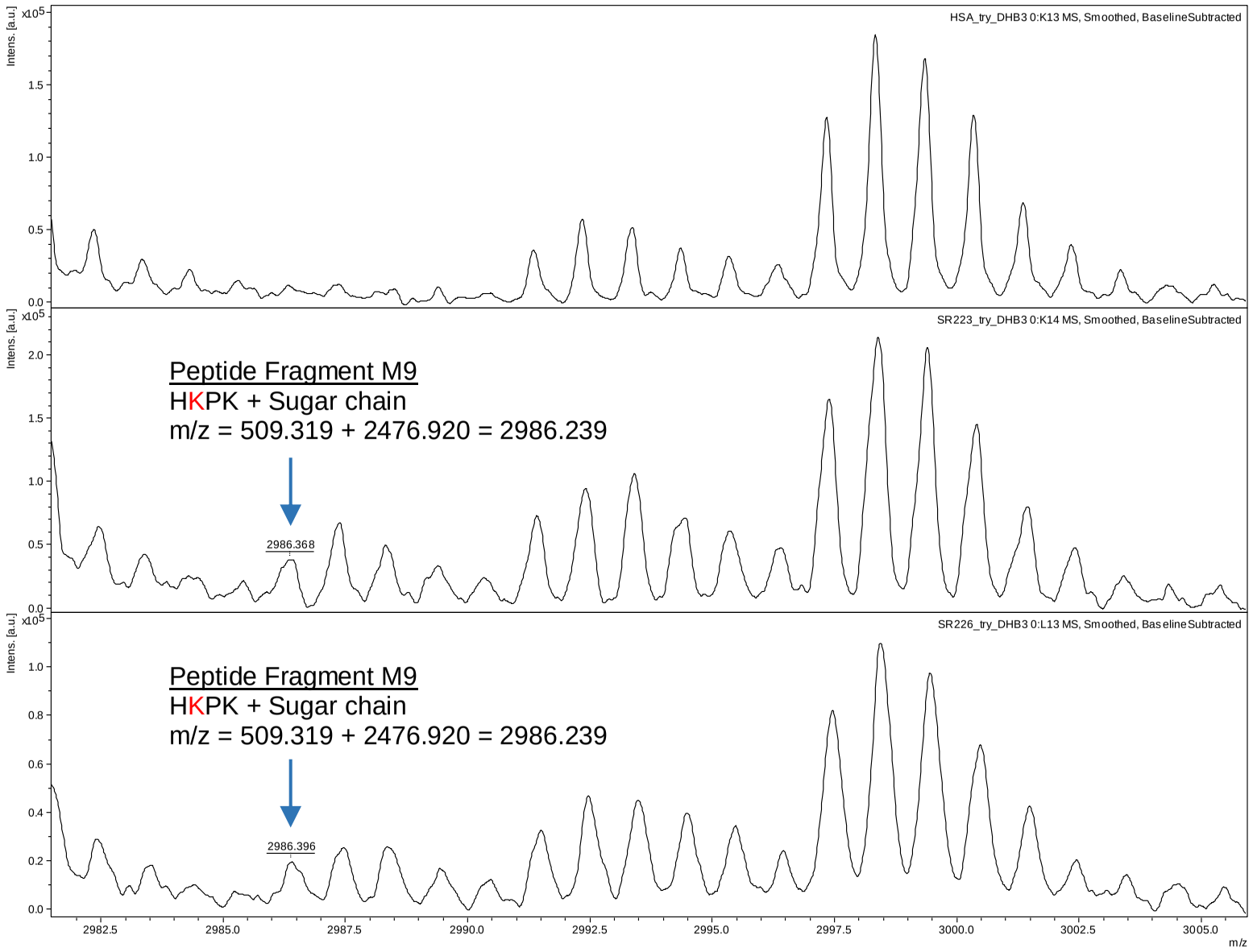


Figure S31. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M9 [HKPK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

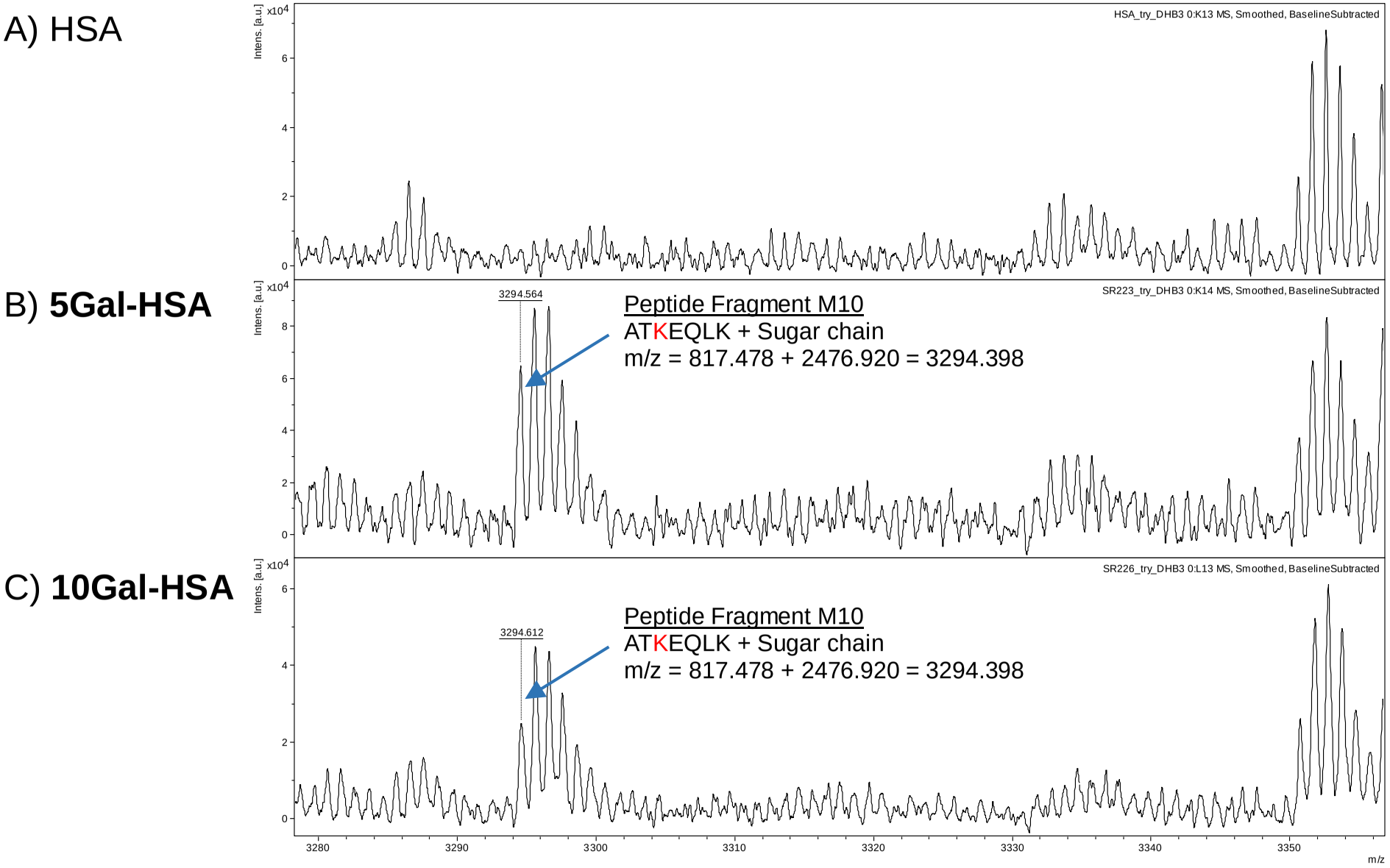


Figure S32. Magnified MALDI-TOF spectrum profile focusing on peptide fragment M10 [ATKEQLK] in protein samples for A) HSA, B) **5Gal-HSA**, and C) **10Gal-HSA**.

2) Cell-based Experiments

2.1 General Cell Culture Protocol

The eleven cell lines used in this study were obtained from either American Type Culture Collection (Virginia, USA), RIKEN Cell Bank, or JCRB Cell Bank. Specific growth media used for each individual cell lines are indicated as follows.

Name	Type	Medium	FBS	Penicillin-Streptomycin	Supplement
A549	human adenocarcinomic alveolar basal epithelial cells	DMEM	10%	1%	
HuH-7	human liver carcinoma cells	DMEM	10%	1%	
AR42J	rat pancreatic cancer cells	DMEM	10%	1%	
HeLa229	human cervical cancer cells	DMEM	10%	1%	
Hep-2	human epithelial type 2 cells	MEM	10%	1%	
U87MG	human neuronal glioblastoma cells	MEM	10%	1%	
DLD-1	human colon adenocarcinoma cells	RPMI1640	10%	1%	
OVCAR-3	human ovary adenocarcinoma cells	RPMI1640	10%	1%	
RL95-2	non-polar human uterine epithelial cells	DMEM:F12 (1:1)	10%	1%	0.005 mg/ml insulin
SW620	human colon cancer cells	L-15	10%	1%	
SK-OV-3	human ovarian carcinoma cells	Macoy's 5a	10%	1%	

2.2 Imaging Studies

Cells under study were plated onto 8-well chamber slides at a density of 3×10^3 cells per 300 µl of media. Cells were then grown for approximately 2 days at 37°C. Stock solutions of glycoalbumins **2a-f** were prepared at a concentration of 10 µM. To initiate incubation, media in each well was removed by suction, followed by the addition of 3 µl of each respective glycoalbumin in 300 µl of media (Final concentration of 100 nm per well). As a negative control, solution containing only **TAMRA-HSA** was also incubated at similar concentrations. Cells were then incubated overnight at 37°C. To initiate imaging, cells were washed 3× with PBS buffer (400 µl, 4°C). Cells were then fixated by addition of paraformaldehyde (500 µl, 4°C) and incubated for 5 min at room temperature. Remaining media was then removed by suction, followed by the addition of 500 µl of the nuclear staining agent DAPI (500× diluted solution, 4°C). Following another incubation period of 5 min at room temperature, cells were washed 1× with PBS buffer (400 µl), and prepared for microscopy imaging using a Fluorescent Microscope BX51 (Olympus Corp, Tokyo, Japan). For TAMRA-dye observation, fluorescence was measured at λ_{EX}=530-550nm/λ_{EM}=575nm, while for DAPI-dye observation, fluorescence was measured at λ_{EX}=330-385nm/λ_{EM}=420nm. Images were obtained at 200× magnification. Recorded fluorescence values were then averaged and subtracted from the **TAMRA-HSA** negative control.

3) Animal-based Experiments

3.1 Excretion and Biodistribution Studies

Glycoalbumins **2b**, **2c**, and **2e**, labeled with the near-infrared fluorescent probe (HiLyte Fluor 750®), were prepared at concentrations of 0.25nmol/50µl (low dosage) and 1.5nmol/30µl (high dosage). Samples were then diluted in 50µL and 70µL saline, respectively, in order to reach a total volume of 100 µL. Solutions were injected into 8 to 10 week-old BALB/cAJcl-nu/nu mice via the tail vein (n=4). Mice were then anesthetized with pentobarbital or isoflurane and placed in a fluorescence imager, IVIS kinetics fluorescence imager® (Caliper Life Sciences, Massachusetts, USA). Dorsal side images were collected at 30 min intervals. Given that tumor tissues were implanted on mouse backs, only dorsal data allowed for the proper monitoring of tumor adhesion.

All procedures involving experiment animals were approved by the Ethics Committee of RIKEN (MAH21-19-17). The experiments were performed in accordance with the institutional and national guidelines.

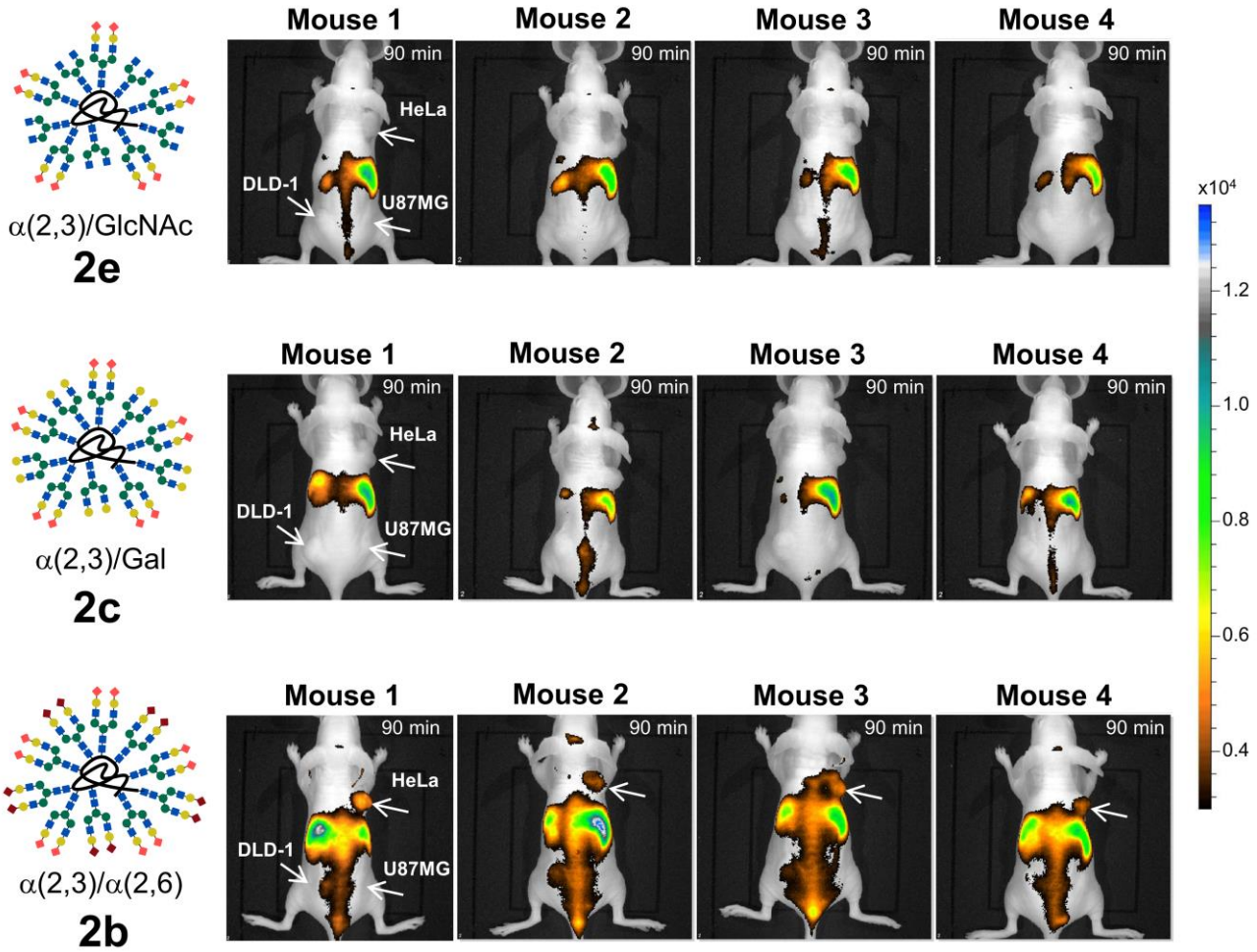


Figure S33. Replicate data (n=4) for noninvasive fluorescence imaging of BALB/c nude mice (dorsal view) at 90 minutes following the intravenous injection of B) α(2,3)Sia/GlcNAc terminated glycoalbumin **2e**, B) α(2,3)Sia/Gal terminated glycoalbumin **2c**, and C) α(2,3)Sia/α(2,6)Sia terminated glycoalbumin **2b**. Tumor locations are as listed; HeLa229 at right shoulder, U87MG at right groin, and DLD-1 at left groin.

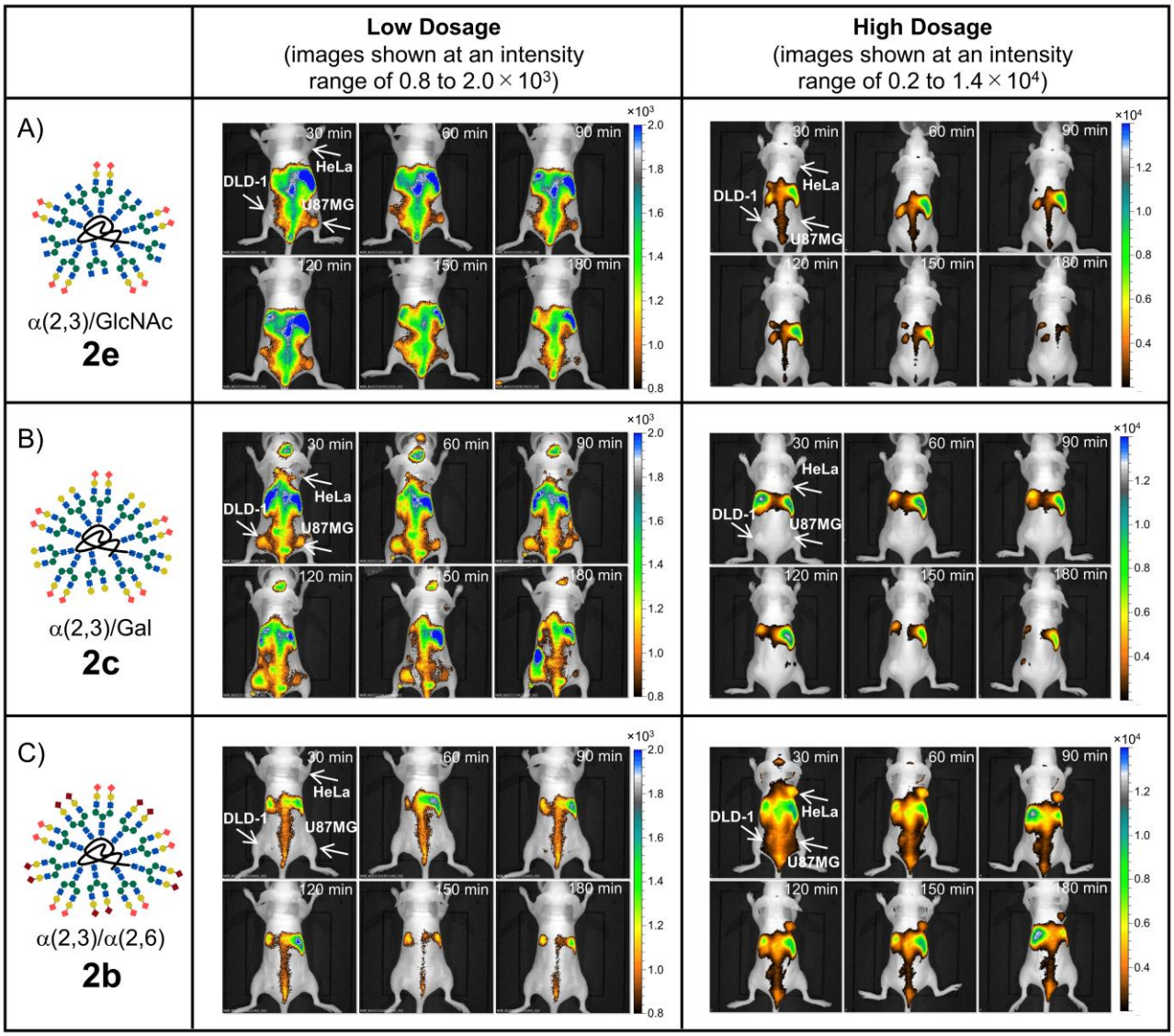


Figure S34. Noninvasive fluorescence imaging of BALB/c nude mice (dorsal view) to compare the differing levels of tumor tissue accumulation between low and high dosages of A) α(2,3)Sia/GlcNAc terminated glycoalbumin **2e**, B) α(2,3)Sia/Gal terminated glycoalbumin **2c**, and C) α(2,3)Sia/α(2,6)Sia terminated glycoalbumin **2b**. The intensity range at which fluorescence was detected at are as indicated. Tumor locations are as listed; HeLa229 at right shoulder, U87MG at right groin, and DLD-1 at left groin.

4) References

1. A. Ogura, T. Tahara, S. Nozaki, K. Morimoto, Y. Kizuka, S. Kitazume, M. Hara, S. Kojima, H. Onoe, A. Kurbangalieva, N. Taniguchi, Y. Watanabe, K. Tanaka, *Sci. Rep*. **2016**, *6*, 21797