

Supplementary Information

Design and Synthesis of Glyco-peptides as Anti-Cancer Agents Targeting Thrombin- Protease Activated Receptor-1 Interaction

Yu-Hsuan Chang^{a,b}, Jen-Chine Wu^c, Hui-Ming Yu^a, Hua-Ting Hsu^a, Ying-Ta Wu^a, Alice
Lin-Tsing Yu^{a,c}, Cheng-Der Tony Yu^d and Chi-Huey Wong^{a,b}.

^a The Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan

^b Department of Chemistry, National Taiwan University, Taipei 106, Taiwan

^c Institute of Stem Cell and Translational Cancer Research, Chang Gung Memorial
Hospital at Linkou, Taoyuan 333423, Taiwan

^d OBI Pharma, Inc., Taipei 11561, Taiwan.

Table of Content

Materials and Methods

Molecular modeling and computed binding energies of designed compounds..	S-3
Chemical synthesis of <i>O</i> -linked glycans.....	S-5
Solid-phase peptide or glycopeptide synthesis.....	S-8
Design of <i>N</i> -terminal PAR-1 sequence as thrombin substrate.....	S-9
Enzyme assays.....	S-10
Flow cytometry, Cell lines and mice	S-
11	
Anticancer activity of HS compounds.....	S-12
References.....	S-13
NMR Spectral data.....	S-14
Mass Spectral data.....	S-17

Materials and Methods

Molecular modeling and computed binding energies of designed compounds

The design of HS peptides was based on the C-terminal region of hirudin binding to thrombin's exosite I, the binding site of PAR-1 to thrombin. The 3D structures of hirudin-thrombin and PAR1-thrombin complexes (PDB codes 4HTC and 3LU9) were used as templates to build the initial structures for the modeling of these designed peptides. The molecular modeling was performed with Schrodinger Suit Program (Schrodinger, Portland). Briefly, molecular docking (using Glide) followed by molecular dynamic (using MacroModel) were applied to refine the interacting pose of each HS peptide in the binding pocket. The OPLS_2005 forcefield was used and water was included in the modeling. In the end, MacroModel Embrace of the Suit was applied to compute the relative binding affinity of each HS peptide with thrombin. The Embrace common parameters and PRCG method with 50000 iterations and 0.05 gradient convergence threshold were used in the computation. In the complex refinement, the receptor residues within 5A of the ligand and the ligand atoms, except the backbone atoms, were freely to move initially followed by no constrain to all atoms of the ligand and the receptor residues that were within 8A of the ligand.

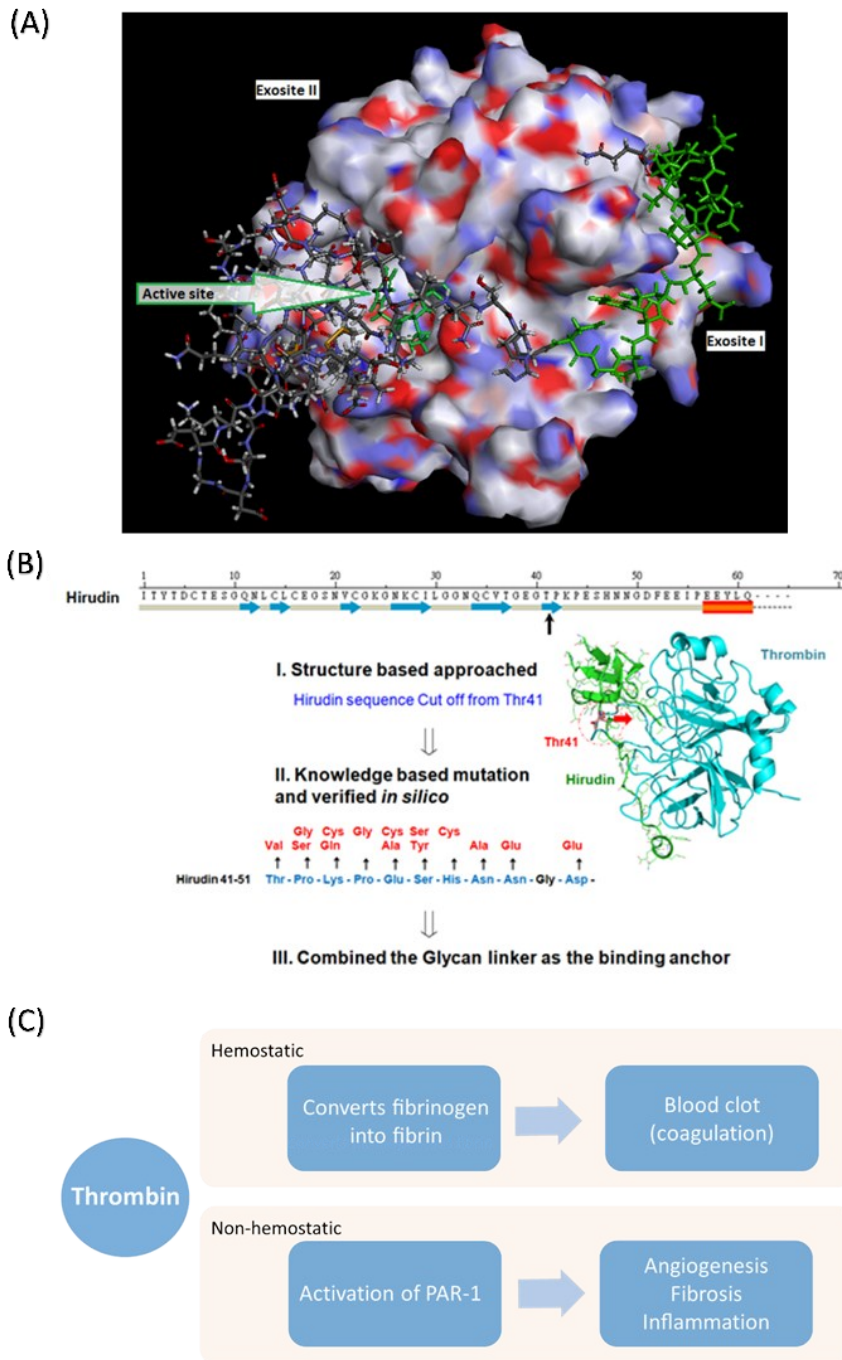


Figure S1. (A) The 3D structure of human α -thrombin binding to hirudin. The three major domains of human α -thrombin (PDB: 4HTC) were highlighted and hirudin was shown to bind to the exosite I and the active site; (B) flow chart of the HS compound design. hirudin was used as a template and reduced its length to identify Thr-41 as the cutoff residue. Then, we modified the amino acids by design using the homology modeling method and built up the 3D structure of the designed peptide and through binding energy studies to optimize the designed peptides; (C) schematic presentation of thrombin-mediated hemostatic and non-hemostatic effects.

Table S1. The calculated relative binding affinity of each complex model of HS peptides and glycopeptides with α -thrombin.

	Del Total Energy (KJ/mol)	Del Solvation Energy (KJ/mol)	Del vdW Energy (KJ/mol)	Del Electrostatic Energy (KJ/mol)
HS-6	-343	2897	-369	-2871
HS-7	-350	2805	-397	-2757
HS-12	-357	4114	-432	-4039
HS-13	-386	4110	-480	-4016
HS-14	-405	4939	-415	-4930
HS-15	-360	4862	-428	-4794
HS-16	-423	4946	-417	-4971
HS-18	-317	3898	-424	-3791
HS-20	-380	4004	-458	-3926
HS-21	-338	4538	-412	-4464
HS-22	-321	3989	-297	-4013
HS-23	-384	4900	-417	-4867
HS-24	-389	4915	-450	-4854
HS-25	-387	4861	-440	-4808
HS-26	-380	4876	-434	-4822
HS-40	-406	4951	-406	-4950
HS-41	-422	4946	-411	-4958

Chemical synthesis of *O*-linked glycans

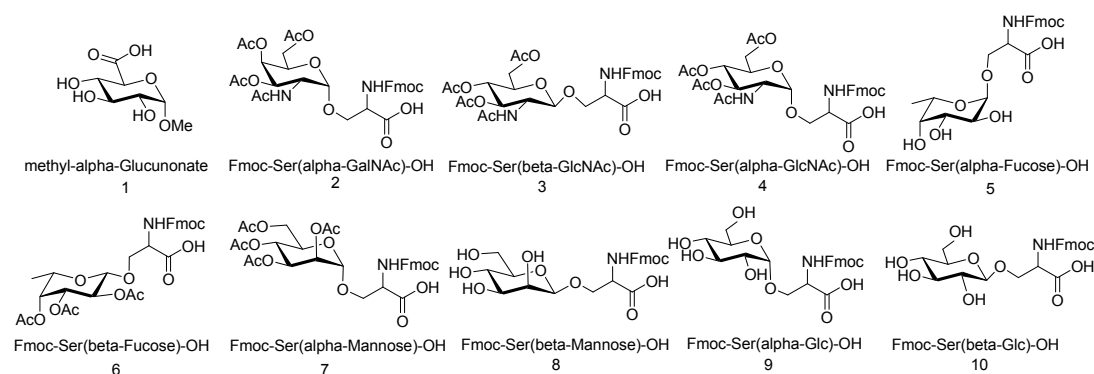
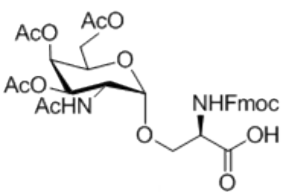
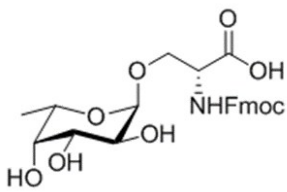
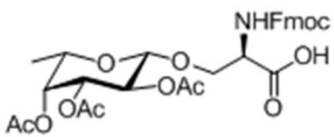


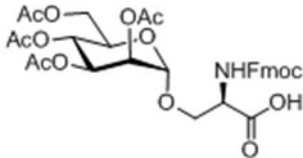
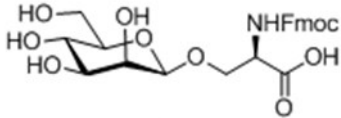
Figure S2. The structures of N-terminal and *O*-linked glycans.

Reagent-grade chemicals were used without further purification and were purchased from Acros, Merck and Sigma-Aldrich. Reactions were monitored by thin-layer chromatography (TLC) on silica gel 60 F-254 plate (2mm, Merck) and visualized under

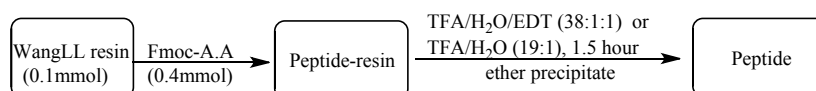
UV illumination and stained with acid ceric ammonium molybdate or *p*-anisaldehyde. Flash column chromatography was performed on silica gel (40–63 μ m, Merck) or LiChroprep RP-18 (40–63 μ m, Merck). NMR spectra were recorded at 600 MHz (^1H NMR) spectrometers in a Bruker Advance 600. The chemical shift was reported in ppm (δ scale) and referenced to residual solvent peaks: chloroform (CDCl_3 : ^1H δ 7.24), deuterated water (D_2O : ^1H δ 4.80) or deuterated methanol (MeOD: ^1H δ 4.78). The coupling constant in Hz was calculated from the chemical shift of ^1H NMR spectra. Data are expressed as follows: chemical shift, multiplicity (s = singlet, d= doublet, t = triplet, q = quartet, m= multiplet), integration and coupling constant (*J*) in Hz. Compounds **1**, **9** and **10** were synthesized and provided by Dr. C.T. Ren. Compounds **3** and **4** were synthesized and provided by Dr. C.C. Lin. Other known compounds were cited with references^{1,2,3}. The structures obtained were characterized by ^1H NMR (Table S2).

Table S2: Characterization data of *O*-linked glycans

Code	Chemical structure	^1H NMR (600 MHz)
2		^1H NMR (600 MHz, MeOD): δ 7.67 (d, <i>J</i> = 7.5 Hz, 2H), 7.55 (dd, <i>J</i> = 11.8, 7.6 Hz, 2H), 7.27 (t, <i>J</i> = 7.4 Hz, 2H), 7.21 (t, <i>J</i> = 7.2 Hz, 2H), 5.28 (d, <i>J</i> = 2.0 Hz, 1H), 5.07 (dd, <i>J</i> = 11.4, 2.3 Hz, 1H), 4.32 (dd, <i>J</i> = 11.4, 4.1 Hz, 2H), 4.24 (dd, <i>J</i> = 18.7, 8.2 Hz, 1H), 4.20 (s, 1H), 4.15 – 4.09 (m, 2H), 3.96 – 3.89 (m, 2H), 3.85 (d, <i>J</i> = 7.5 Hz, 1H), 3.82 – 3.78 (m, 1H), 2.01 (s, 3H), 1.84 (d, <i>J</i> = 7.9 Hz, 6H), 1.78 (s, 3H).
5		^1H NMR (600 MHz, MeOD): δ 7.70 (d, <i>J</i> = 7.5 Hz, 2H), 7.63 – 7.53 (m, 2H), 7.30 (t, <i>J</i> = 7.5 Hz, 2H), 7.25 – 7.20 (m, 2H), 4.64 (d, <i>J</i> = 3.5 Hz, 1H), 4.33 – 4.24 (m, 2H), 4.14 (dd, <i>J</i> = 7.6, 5.4 Hz, 2H), 4.03 (dd, <i>J</i> = 9.2, 3.7 Hz, 1H), 3.87 (q, <i>J</i> = 6.3 Hz, 1H), 3.65 (dd, <i>J</i> = 10.1, 3.2 Hz, 1H), 3.58 (dd, <i>J</i> = 10.1, 3.7 Hz, 1H), 3.51 (d, <i>J</i> = 2.5 Hz, 1H), 3.44 (dd, <i>J</i> = 9.3, 2.5 Hz, 1H), 1.10 (d, <i>J</i> = 6.5 Hz, 3H).
6		^1H NMR (600 MHz, MeOD): δ 7.71 (d, <i>J</i> = 7.5 Hz, 2H), 7.59 (d, <i>J</i> = 6.3 Hz, 2H), 7.31 (dd, <i>J</i> = 9.5, 5.1 Hz, 2H), 7.24 (t, <i>J</i> = 7.3 Hz, 2H), 5.11 (d, <i>J</i> = 2.4 Hz, 1H), 4.96 (qd, <i>J</i> = 10.5, 5.3 Hz,

		2H), 4.42 (d, $J = 7.2$ Hz, 1H), 4.34 (dd, $J = 10.4$, 7.0 Hz, 1H), 4.27 (dd, $J = 10.5$, 6.7 Hz, 1H), 4.13 (dd, $J = 13.5$, 5.5 Hz, 2H), 3.97 (dd, $J = 10.3$, 2.7 Hz, 1H), 3.91 (dd, $J = 10.3$, 4.9 Hz, 1H), 3.75 (q, $J = 6.1$ Hz, 1H), 2.04 (s, 3H), 1.92 (s, 3H), 1.85 (s, 3H), 1.01 (d, $J = 6.3$ Hz, 3H).
7		^1H NMR (600 MHz, MeOD): δ 7.71 (d, $J = 7.4$ Hz, 2H), 7.61 (t, $J = 6.7$ Hz, 2H), 7.30 (t, $J = 7.5$ Hz, 2H), 7.23 (t, $J = 7.4$ Hz, 2H), 5.22 (dd, $J = 10.1$, 3.1 Hz, 1H), 5.13 (dd, $J = 19.1$, 8.8 Hz, 2H), 4.40 (q, $J = 9.8$ Hz, 1H), 4.25 (s, 1H), 4.13 (ddd, $J = 17.5$, 13.0, 6.5 Hz, 3H), 4.04 – 3.97 (m, 3H), 3.83 (dd, $J = 10.0$, 5.2 Hz, 1H), 2.03 (s, 3H), 1.94 (s, 3H), 1.84 (d, $J = 11.6$ Hz, 6H).
8		^1H NMR (600 MHz, D_2O): δ 7.89 (t, $J = 7.8$ Hz, 2H), 7.70 (dt, $J = 22.2$, 11.0 Hz, 2H), 7.49 (dd, $J = 16.0$, 8.1 Hz, 2H), 7.42 (dd, $J = 17.5$, 7.6 Hz, 2H), 4.73 – 4.65 (m, 1H), 4.58 – 4.49 (m, 1H), 4.41 (s, 1H), 4.28 (t, $J = 5.0$ Hz, 1H), 4.09 – 4.02 (m, 1H), 3.94 – 3.83 (m, 2H), 3.70 – 3.61 (m, 2H), 3.59 – 3.48 (m, 2H), 3.29 (ddd, $J = 14.3$, 8.2, 4.9 Hz, 1H).

A. Synthesis of peptide



B. Synthesis of glycopeptide

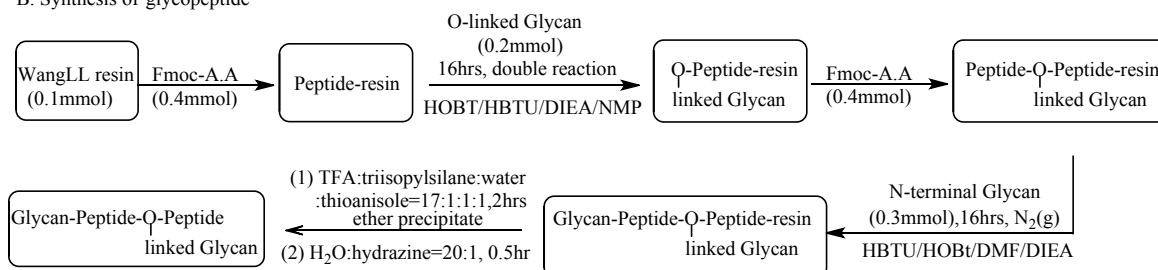


Figure S3. Schematic diagram of solid-phase peptide or glycopeptide synthesis.

The peptides and glycopeptides used in this study were synthesized by Solid Phase Peptide Synthesis (SPPS) using the standard Fmoc chemistry protocol. The Wang resin LL (1 equiv.) was used as solid support for the coupling of Fmoc amino acids (4-6 equiv.) and Fmoc-glycosyl amino acids (2 equiv.). The synthesis was performed on Applied Biosystems 433A peptide synthesizer and Protein Technologies PS3™ Peptide Synthesizer; all reactions were performed at room temperature. The peptide and glycopeptide resins were subjected to a cleavage of cocktail (TFA/thioanisole/triisopropylsilane/water, 17:1:1:1) for 2 h at room temperature. The resin was filtered and the combined filtrates were triturated with cold diethyl ether to give a white suspension, which was centrifuged and the ether layer was subsequently decanted. Then, the precipitate was concentrated in vacuum, and the crude products were purified by HPLC and characterized by MALDI-TOF/MS (matrix α -cyano-4-hydroxycinnamic acid).

Analytical reverse-phase HPLC was performed on the Agilent 1200 series HPLC system. Peptides were analyzed using a SUPELCO™ (Discovery® BIO Wide Pore C18, 25 cm x 10 mm, 5 μ m) column at a flow rate of 1 ml/min using a mobile phase of 2% ACN with 0.1% TFA in water (Solvent A) and 90% ACN with 0.1% TFA in water (Solvent B) with gradients as specified and detected with absorbance detector operating at 214 nm. Results were analyzed with Agilent 1200 series software. All compounds were characterized by NMR and mass spectrometry (See charts at the end).

Design of *N*-terminal PAR-1 sequence as thrombin substrate

Since thrombin will recognize the *N*-terminal extracellular domain of thrombin receptor PAR-1, we synthesized a peptide based on the *N*-terminal ten residues of the PAR-1 exodomain as substrate for thrombin to detect the hydrolytic activity.

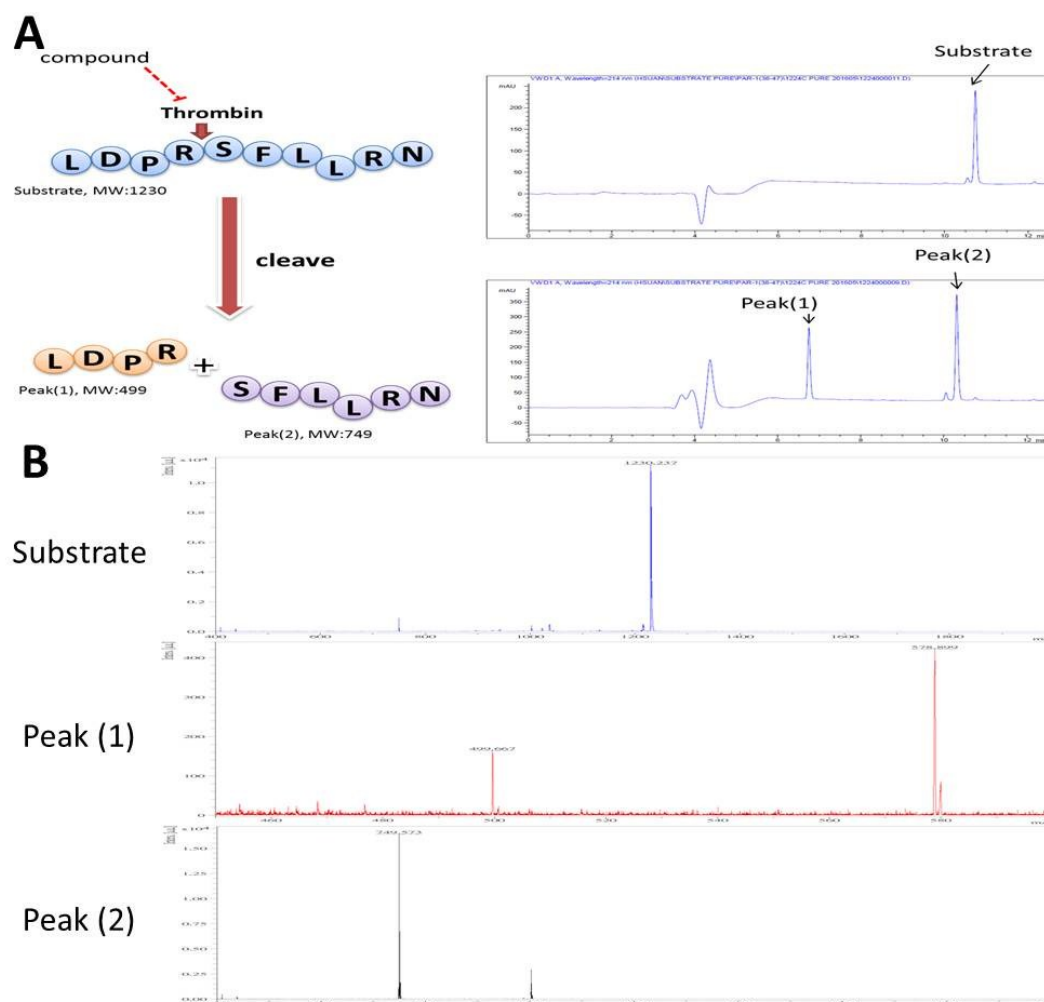


Figure S4. Assay of substrate cleavage in thrombin-PAR-1 interaction. (A) Comparative HPLC analysis and (B) analysis of peptide fragments with mass spectrometry: Substrate 1230 m/z, Peak (1) 499 m/z and Peak (2) 749 m/z.

We monitored the ability of thrombin to cleave the substrate using the HPLC analysis with a gradient from 5% to 50% B solution in 9 min, then with 50% B solution from 9 min to 12.5 min at a flow rate of 1 ml/min. The substrate was detected in 10.6 min and the two cleaved fragments were detected in 6.6 and 10.1 min, respectively.

The substrate assay was relied on detecting the change of signal peak. Using the above assay condition where the substrate (1.72×10^{-3} μmol), thrombin (0.0028×10^{-3} μmol)

and HS compounds (3.44×10^{-3} μmol) were mixed at room temperature for 20 min, the degree of inhibition on the substrate cleavage was monitored. The control group was performed exactly in the same manner, except in the absence of HS compound.

The % of inhibition was calculated from the integral peak (1) area. The control group and the experimental group were used to tentatively estimate the % inhibition using the following formula:

$$\% \text{ Inhibition} = \frac{(\text{Control group integral area} - \text{Experimental group integral area})}{\text{Control group integral area}} \times 100$$

Enzyme assays

Thrombin activity was determined using our designed substrate. Reactions were conducted at 25 °C and initiated by addition of thrombin (0.191 μM final concentration) to each of the six different substrate concentrations in H_2O . Kinetic data were fit to the Lineweaver-Burk plot and the Michaelis-Menten equation: $v_i = V_{\text{max}}[S]/K_m + [S]$, where v_i is the initial velocity, V_{max} is the maximal velocity at saturated concentration of substrate, K_m is the Michaelis constant, and $[S]$ is the substrate concentration

The IC_{50} values for the inhibition of thrombin by hirudin and four HS compounds were determined in the presence of the substrate (57.3 μM). The parameters were determined by varying the concentrations of hirudin (0, 28.67-172 nM), HS-40 (0, 114.67-1834.67 μM), HS-14, 16, 41 (0, 57.3-917.33 μM).

Finally, the K_i values were calculated from Cheng-Prusoff equation: $K_i = \text{IC}_{50}/[1+(S/K_m)]$, where IC_{50} is the concentration of inhibitor that causes 50% inhibition, S is the substrate concentration used in the assay, and K_m is the Michaelis constant.

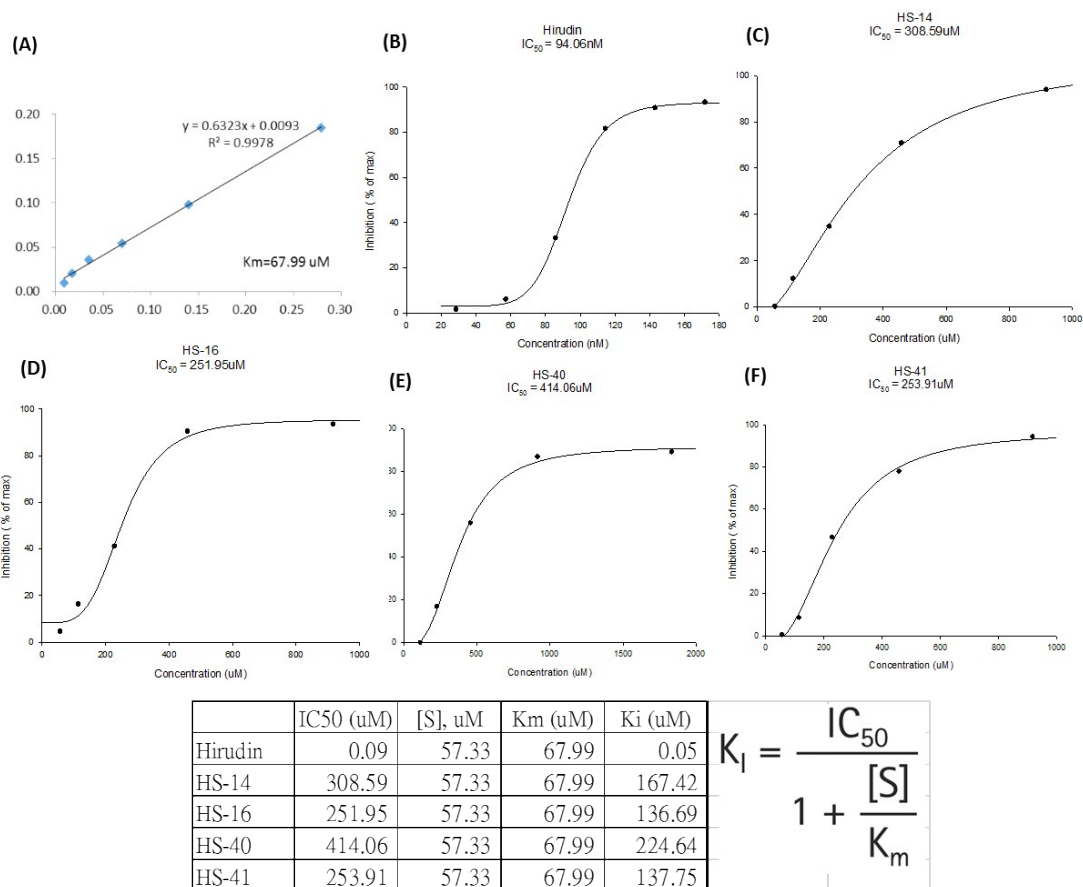


Figure S5. (A) Use of Lineweaver-Burk plot and Michaelis-Menten equation to determine the K_m value of substrate. (B)-(F), determination of the IC_{50} and K_i values using Cheng-Prusoff equation.

Flow cytometry

Cells were grown to 80% confluence and detached from culture dish by EDTA. Then, cells were labeled with anti-PAR-1 antibody (Cyan line; Santa Cruz, H-111) or isotype antibody for 1 h. The excess of antibody was removed by washing with PBS and cells were stained with anti-rabbit antibody-FITC. The expression of PAR-1 was performed on a FACSCanto flow cytometer.

Cell lines and mice

Human non-small lung cancer cells H460 were cultured in Roswell Park Memorial Institute (RPMI)-1640 medium. The media were supplemented with 5% fetal bovine serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin and 100 $\mu\text{g}/\text{ml}$ streptomycin. Cell cultures were incubated in humidified atmosphere of 5% CO_2 at 37 $^\circ\text{C}$.

The NOD/*scid* mice of six to eight weeks old were selected for our studies. All animal experiments conducted in this study were approved by the Institutional Animal Care and Use Committee of the Academia Sinica, Taiwan.

Anticancer activity of HS compounds

Since the designed HS compounds were used to inhibit thrombin and activate the receptor PAR-1 to promote angiogenesis. We set out to identify a proper cell line for further study. The FACSCanto flow cytometer was used to screen several human tumor cell lines, including A549, H460, MIA PaCa-1, MIA PaCa-2, CL1-1 and CL1-5, and finally we identified the PAR-1 positive cell lines CL1-5 (Lung carcinoma), H460 (Lung carcinoma) and MIA PaCa-2 (pancreas carcinoma) as candidates for evaluation of the HS compounds as inhibitors of tumor growth. We then selected H460 human lung cancer cell line to create a model for further animal experiments.

In the lung cancer H460 model, we evaluated the potential of these HS compounds as anti-cancer agents. First, we used pre-treatment protocol as described in the following. We selected PBS and hirudin as control group, and the experimental group includes three types of glycopeptides as inhibitors to investigate their activity: HS-12, HS-18 (peptide), HS-13 (glycopeptide with one glycan) and HS-14, 15, 16 (glycopeptide with two glycans)

The H460 tumor growth assay

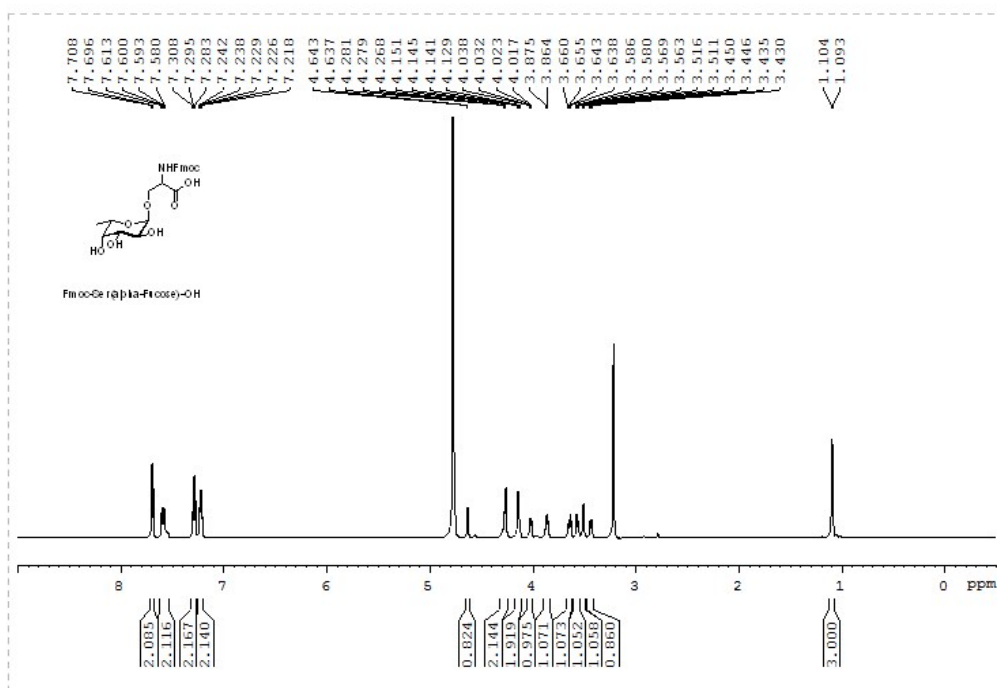
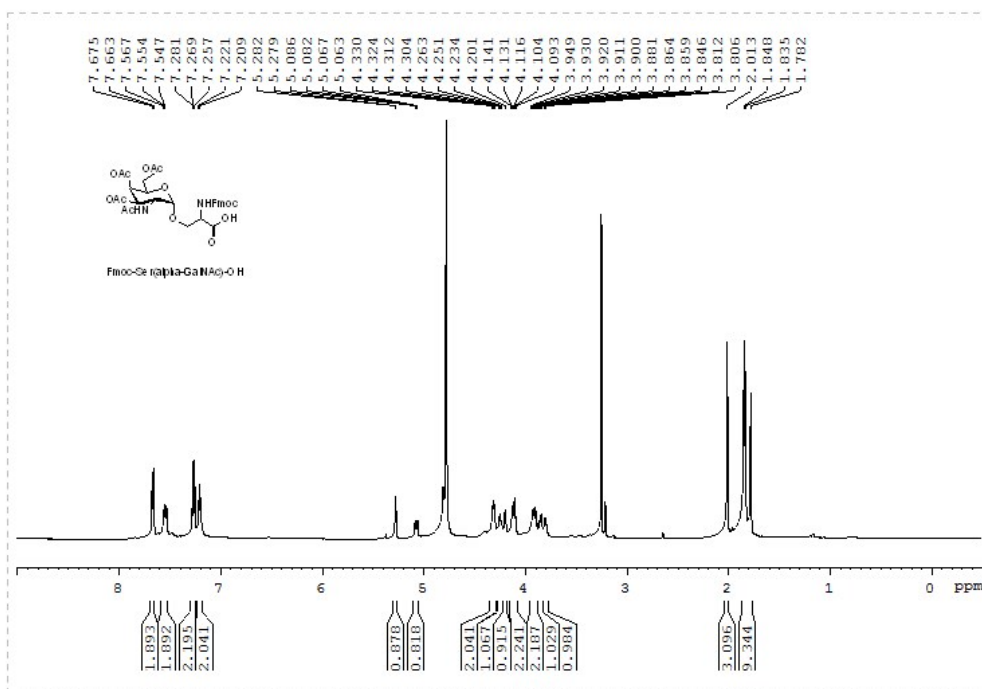
(A) Pre-treated compounds in H460 tumor growth assay. First, the NOD/*scid* mice were treated with hirudin or HS compounds at a dose of 0.036umole or PBS as control via subcutaneous (*s.c.*) injection. After 30min, all mice were inoculated with 1×10^5 cell/mice H460 human lung cancer cells also by subcutaneous injection. Then, after 15 min mice were treated with hirudin or HS compounds at a dose of 0.036umole or PBS as control every other day for consecutive two weeks. These mice were observed and tumor size measured starting on the tenth day. The tumor volumes (mm^3) were calculated using the equation: $AB^2/2$ including the long axis-A (length) and short axis-B (width). Statistical significance between the groups was determined by an unpaired 2-tailed Student's t-test. P value is identified as following: $P < 0.05$, *; $P < 0.01$, **; $P < 0.001$, ***.

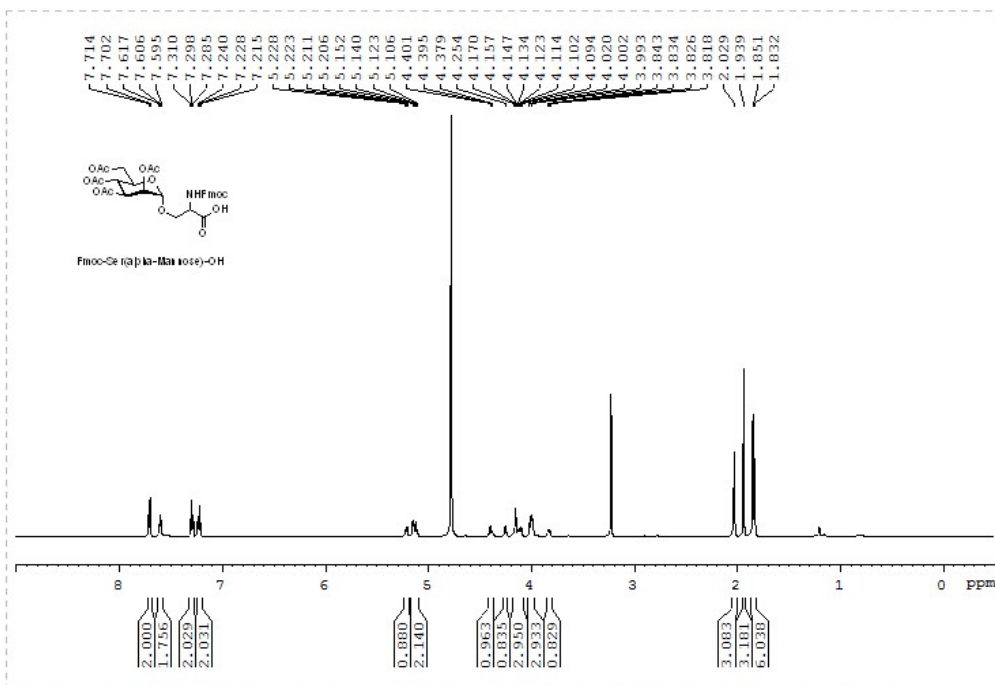
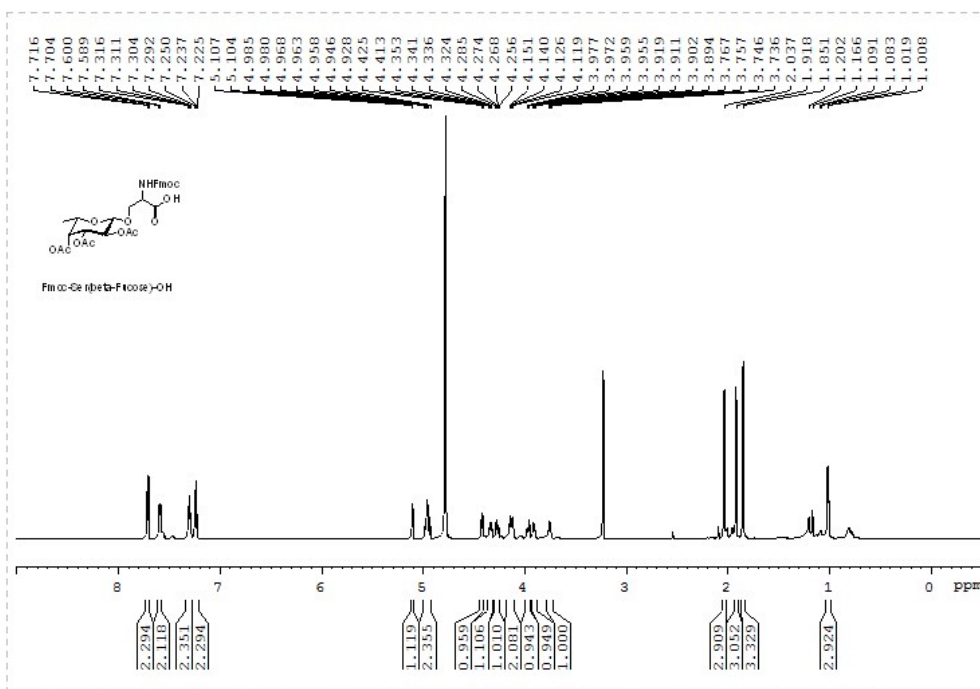
(B) Treated compounds in H460 tumor growth assay. NOD/*scid* mice were subcutaneously injected with 1×10^5 cell/mice H460 human lung cancer cells. After three days, these mice were treated with hirudin or HS compounds at a dose of 0.036 umol or PBS as control via subcutaneous (*s.c.*) injection every other day for

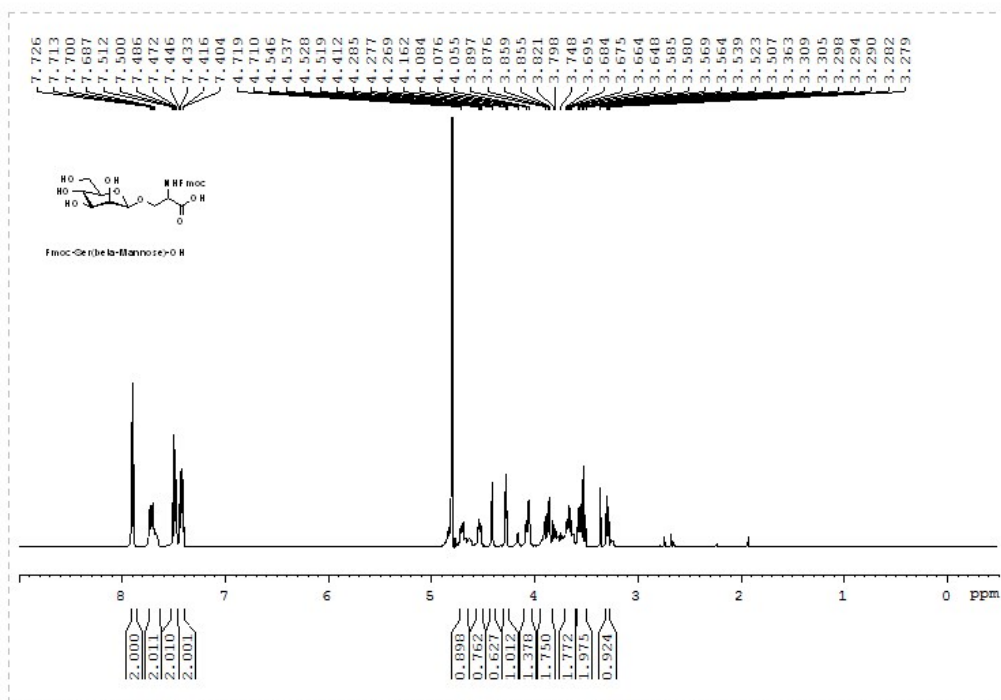
consecutive two weeks. These mice were observed and tumor size measured starting on the sixteenth day. The statistical analysis and measurement of tumor volume were described as above.

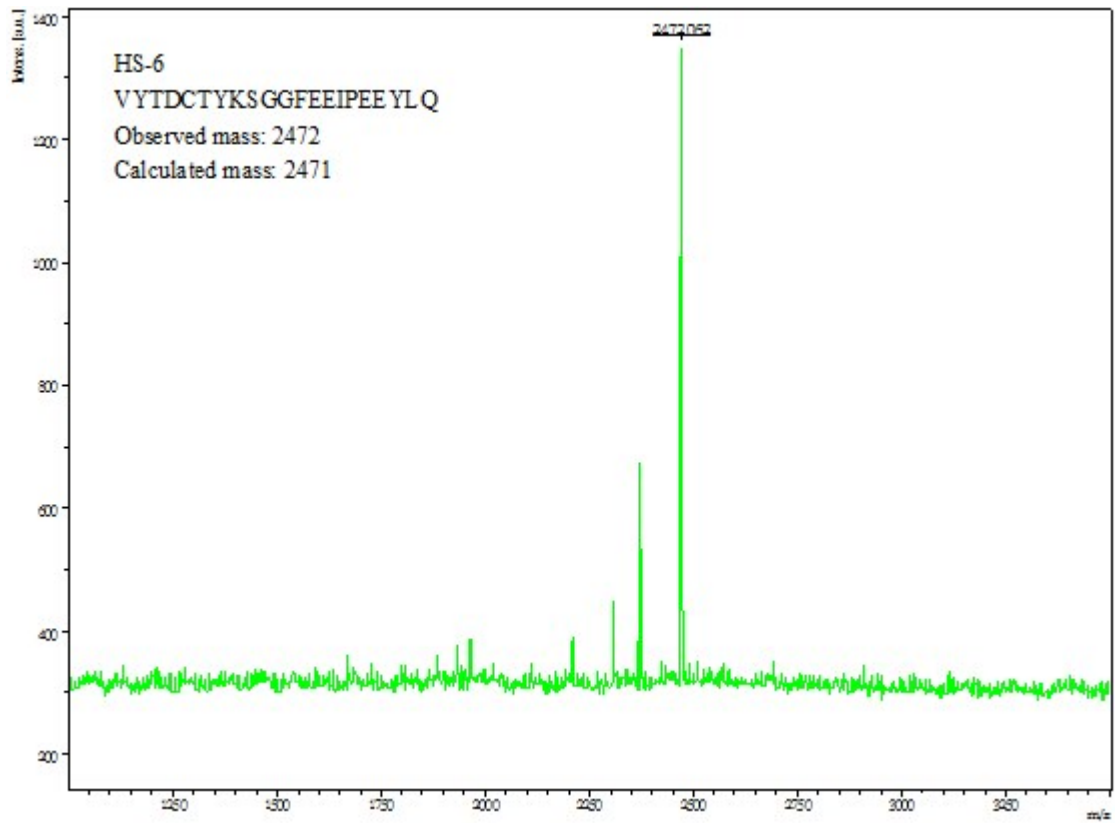
References

- 1 M. Elofsson, S. Roy, L. A. Salvador and J. Kihlberg, *Tetrahedron Lett*, 1996, **37**, 7645.
- 2 L. A. Salvador, M. Elofsson and J. Kihlberg, *Tetrahedron*, 1995, **51**, 5643.
- 3 G. A. Winterfeld, Y. Ito, T. Ogawa and R. R. Schmidt, *Eur J Org Chem*, 1999, **1999**, 1167.

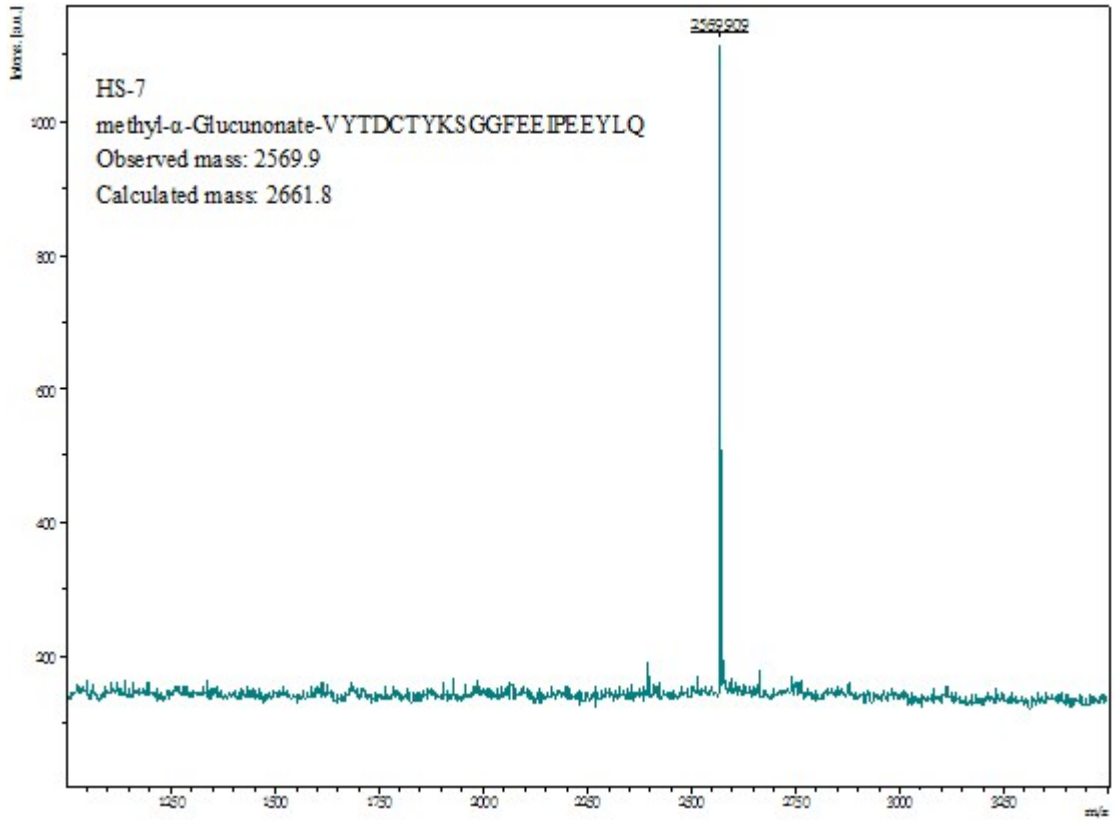




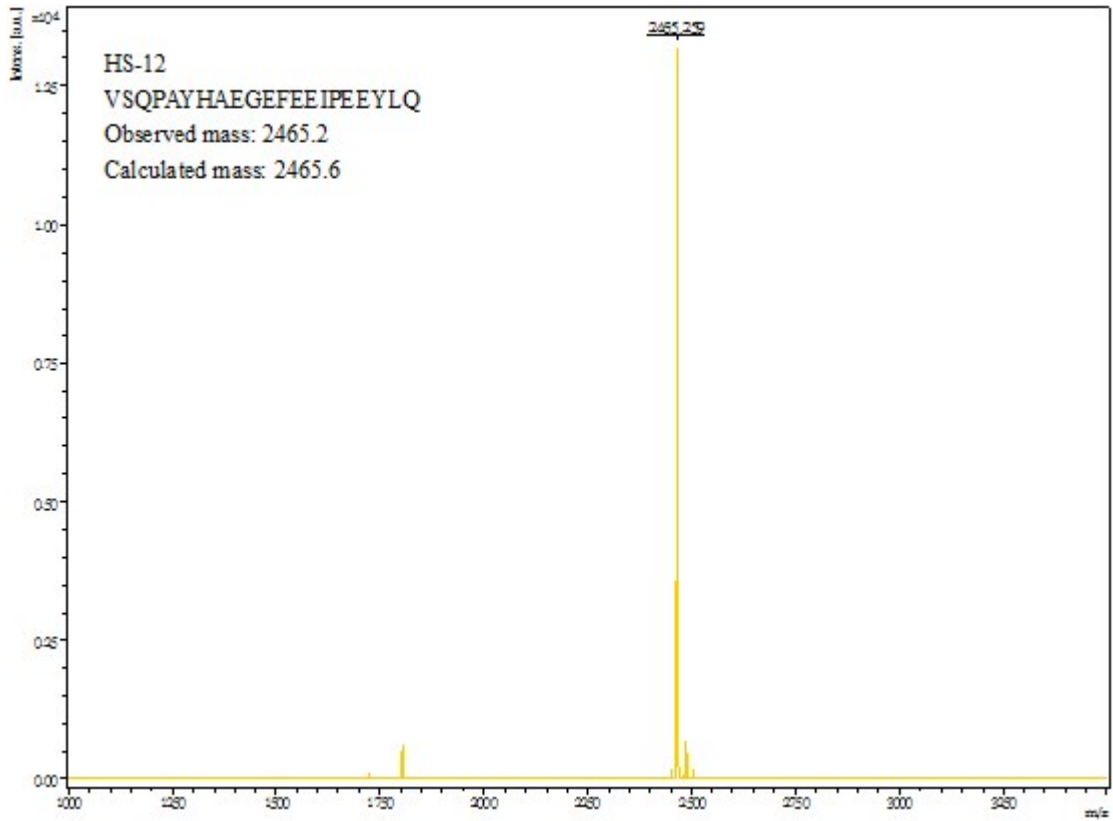




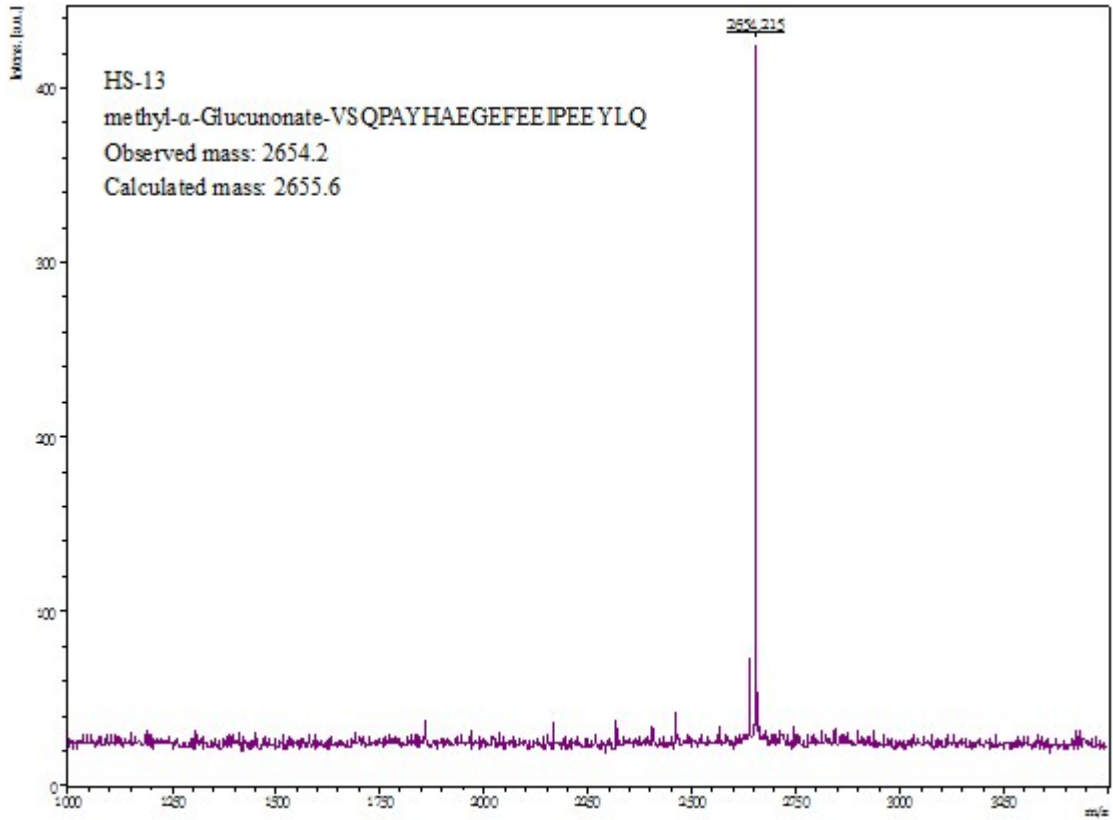
MALDI-TOF mass spectra of HS-6



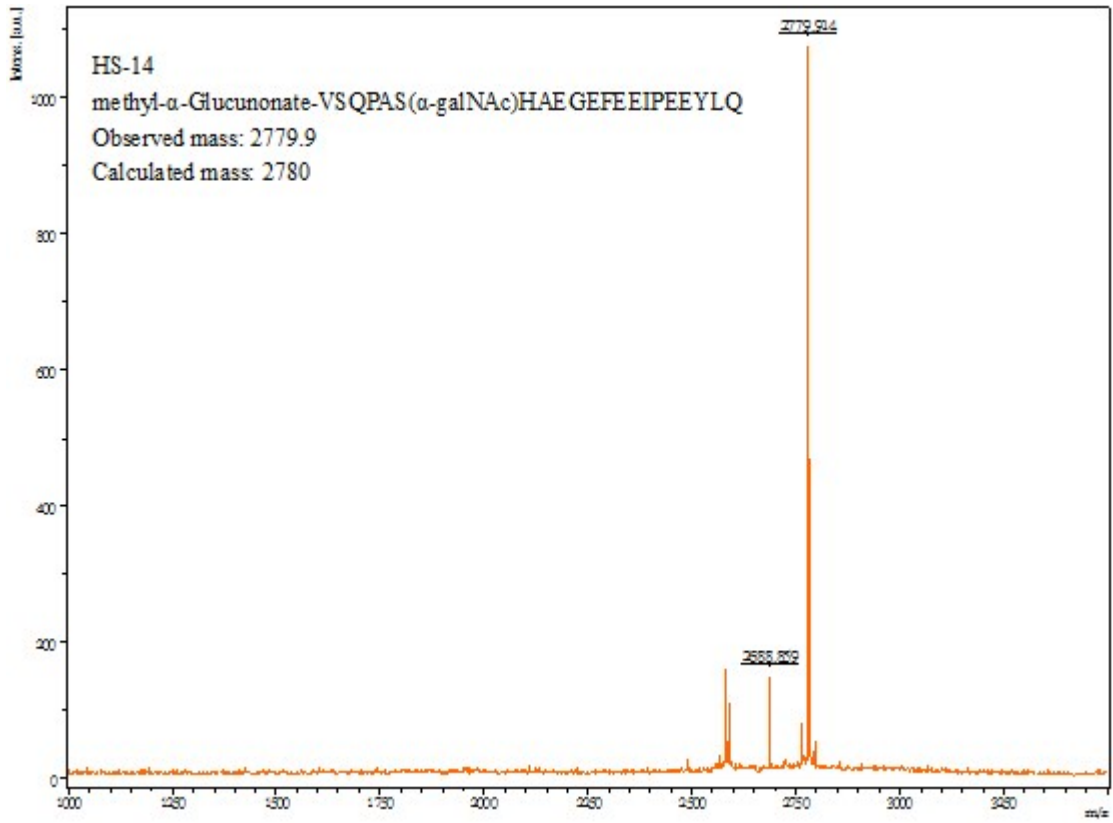
MALDI-TOF mass spectra of HS-7



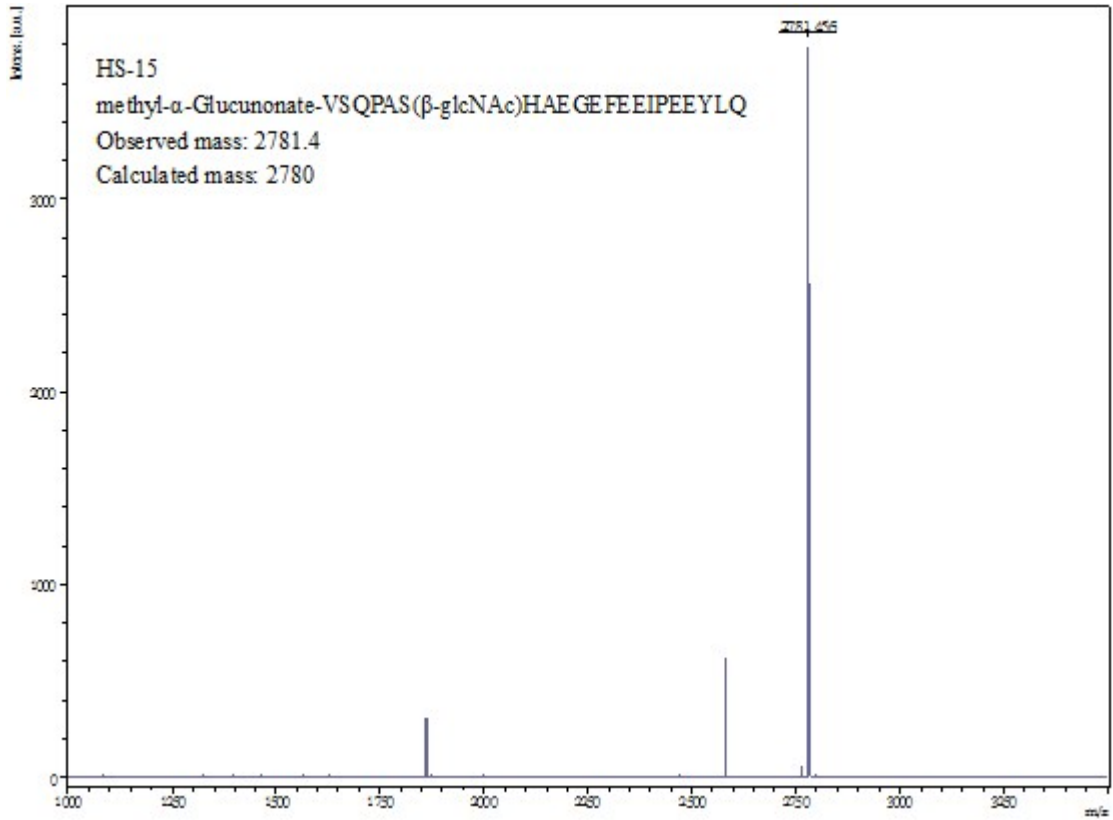
MALDI-TOF mass spectra of HS-12



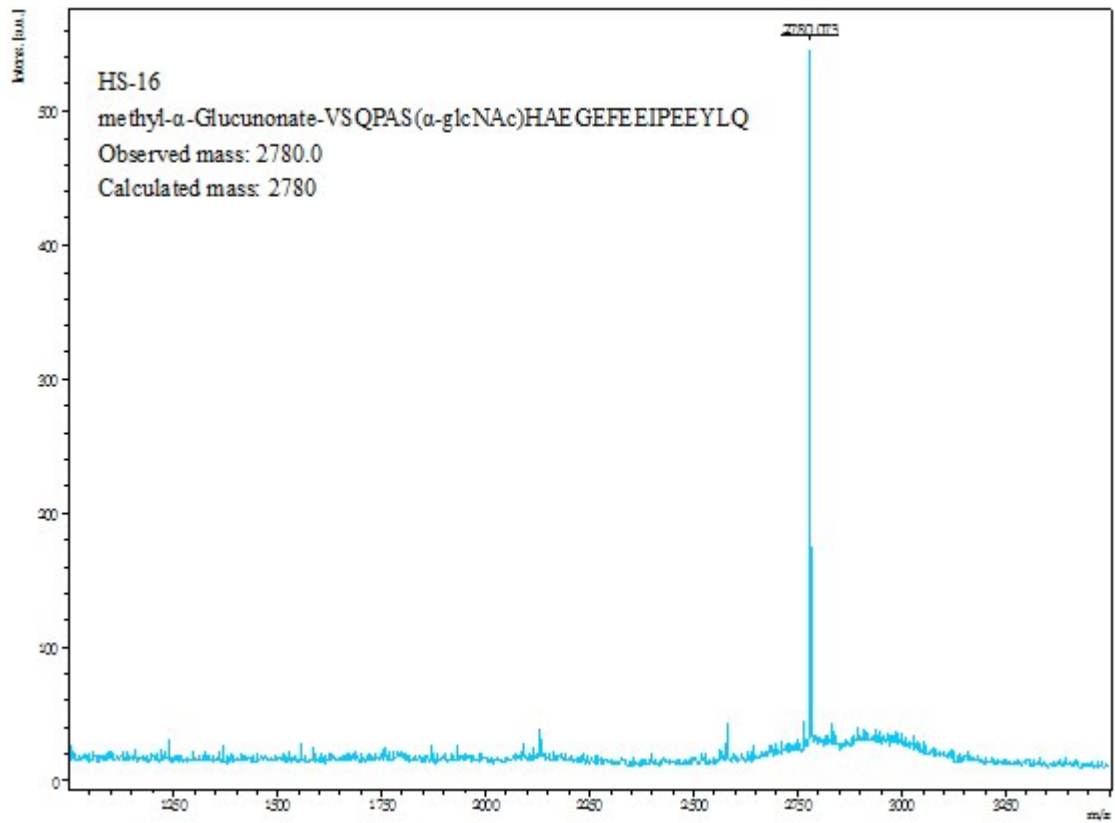
MALDI-TOF mass spectra of HS-13



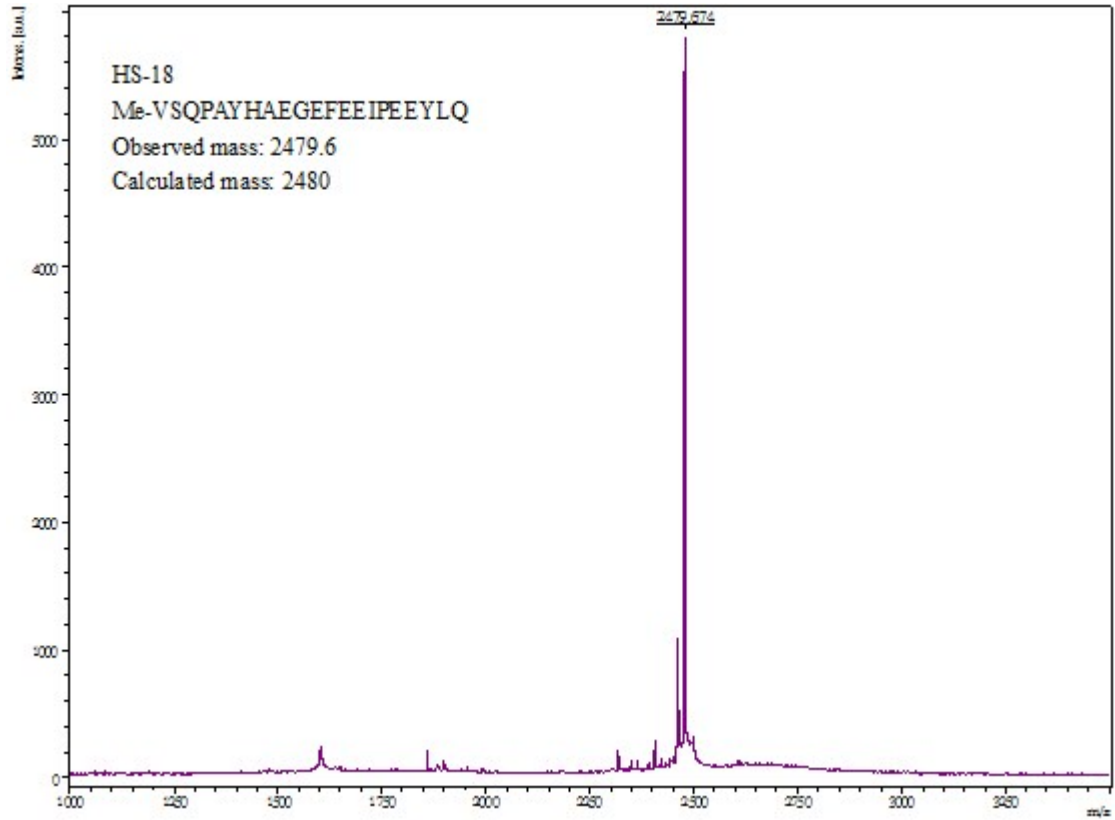
MALDI-TOF mass spectra of HS-14



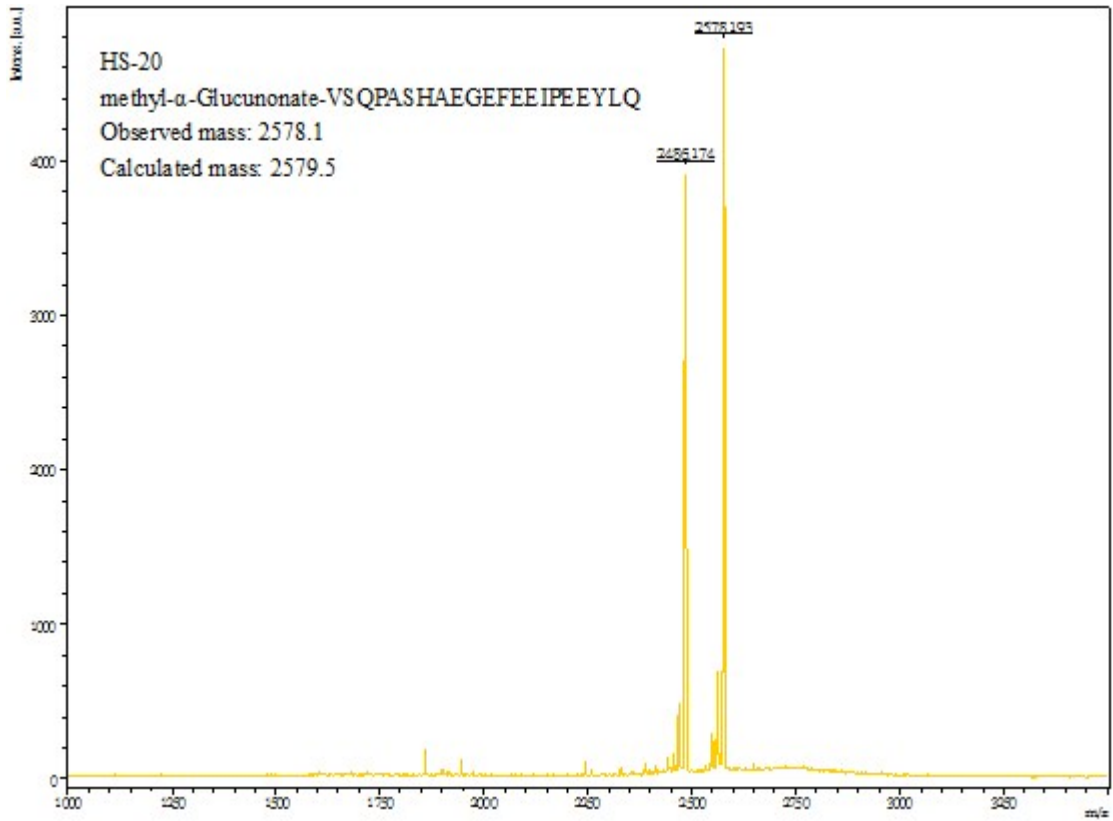
MALDI-TOF mass spectra of HS-15



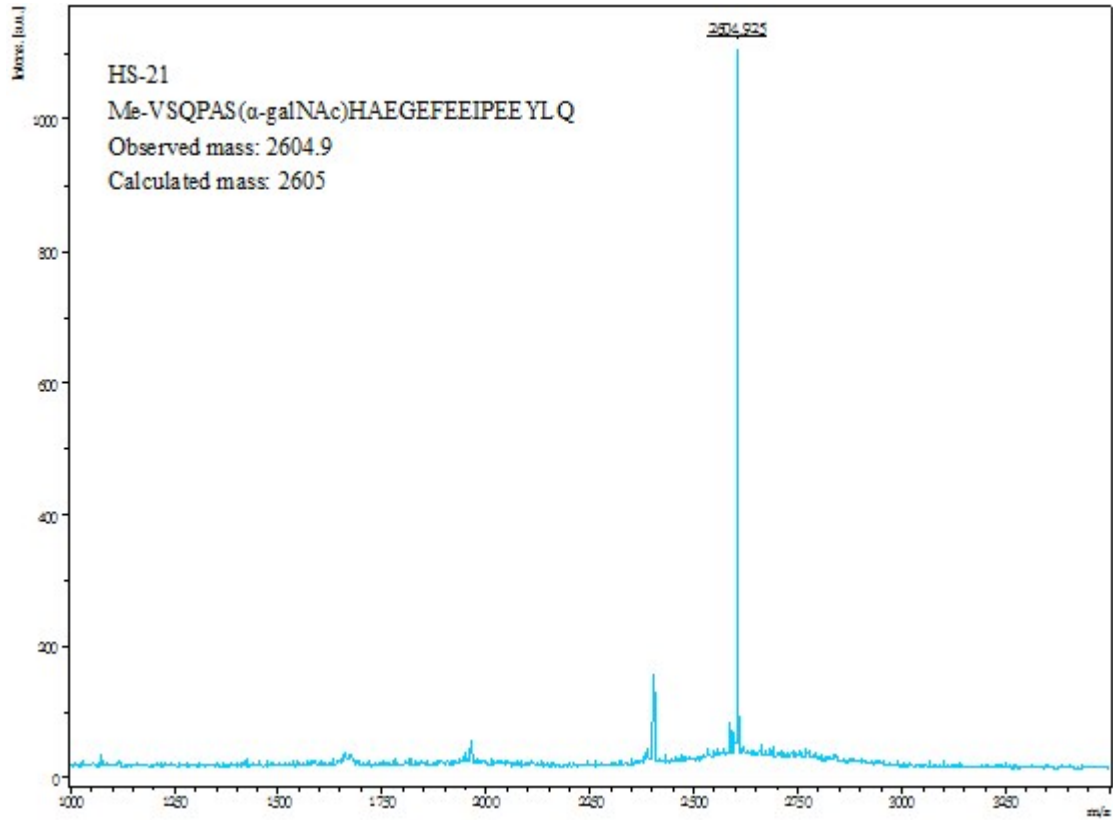
MALDI-TOF mass spectra of HS-16



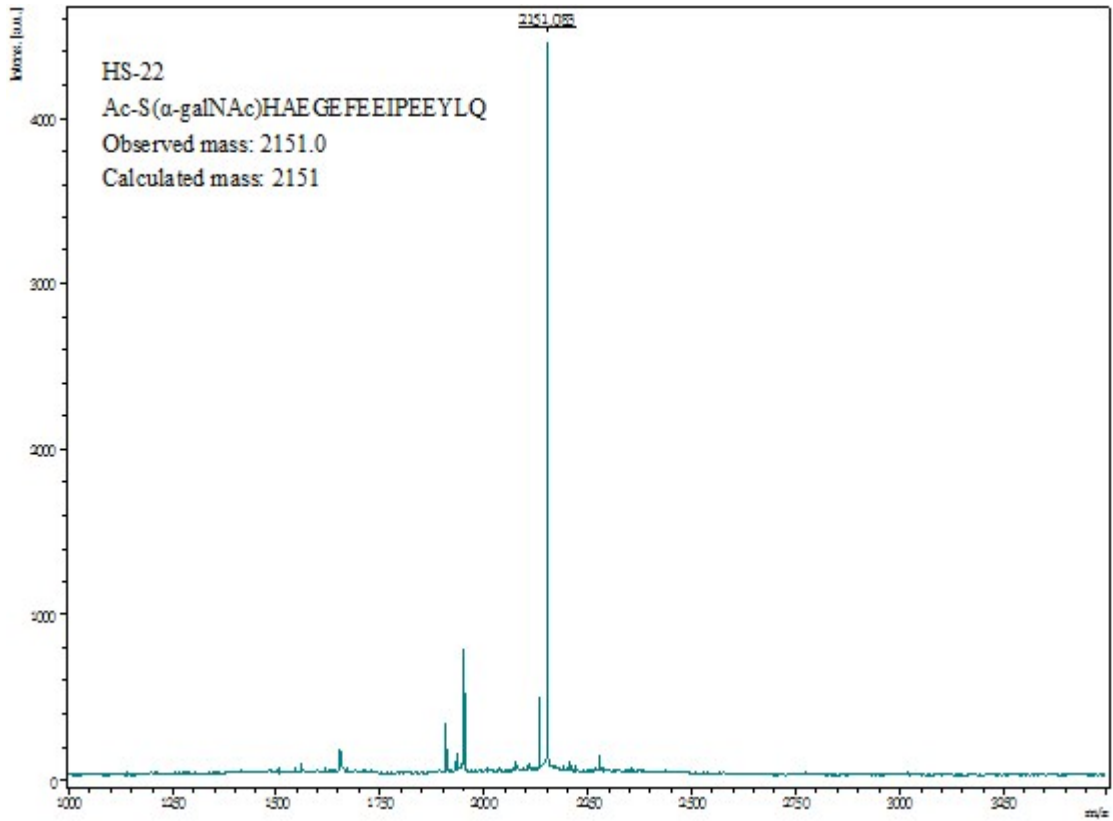
MALDI-TOF mass spectra of HS-18



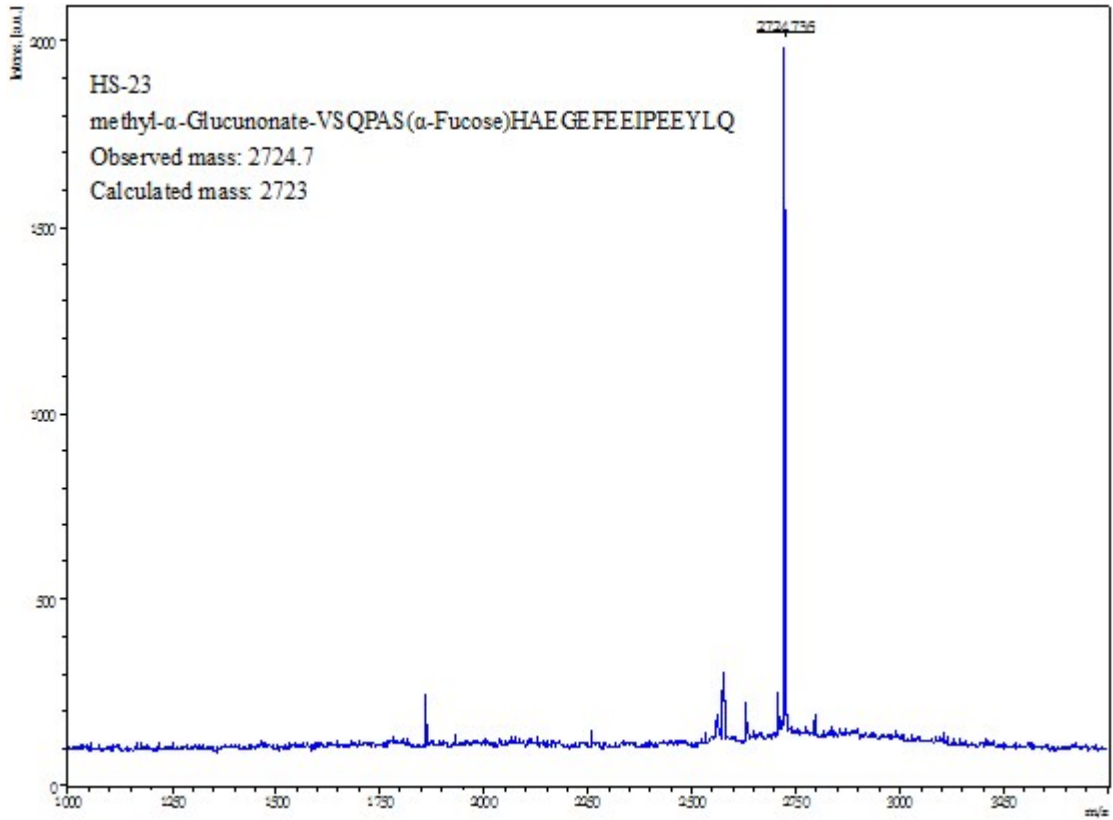
MALDI-TOF mass spectra of HS-20



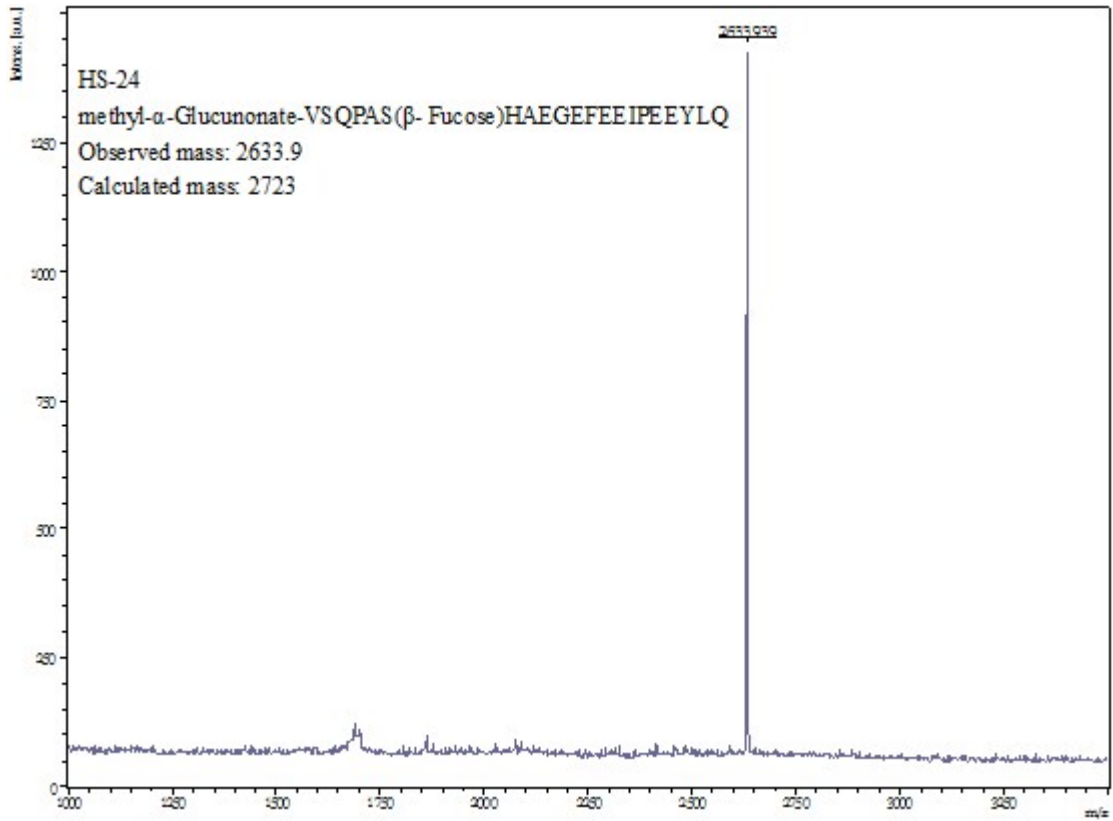
MALDI-TOF mass spectra of HS-21



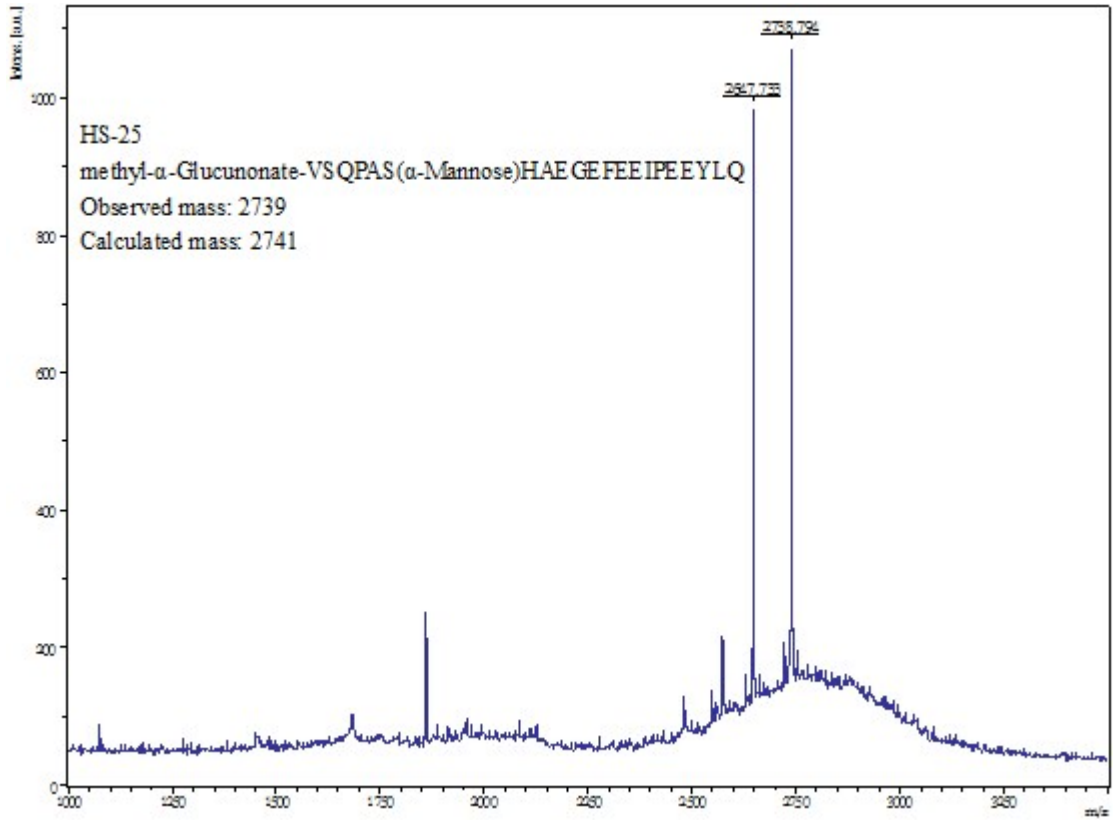
MALDI-TOF mass spectra of HS-22



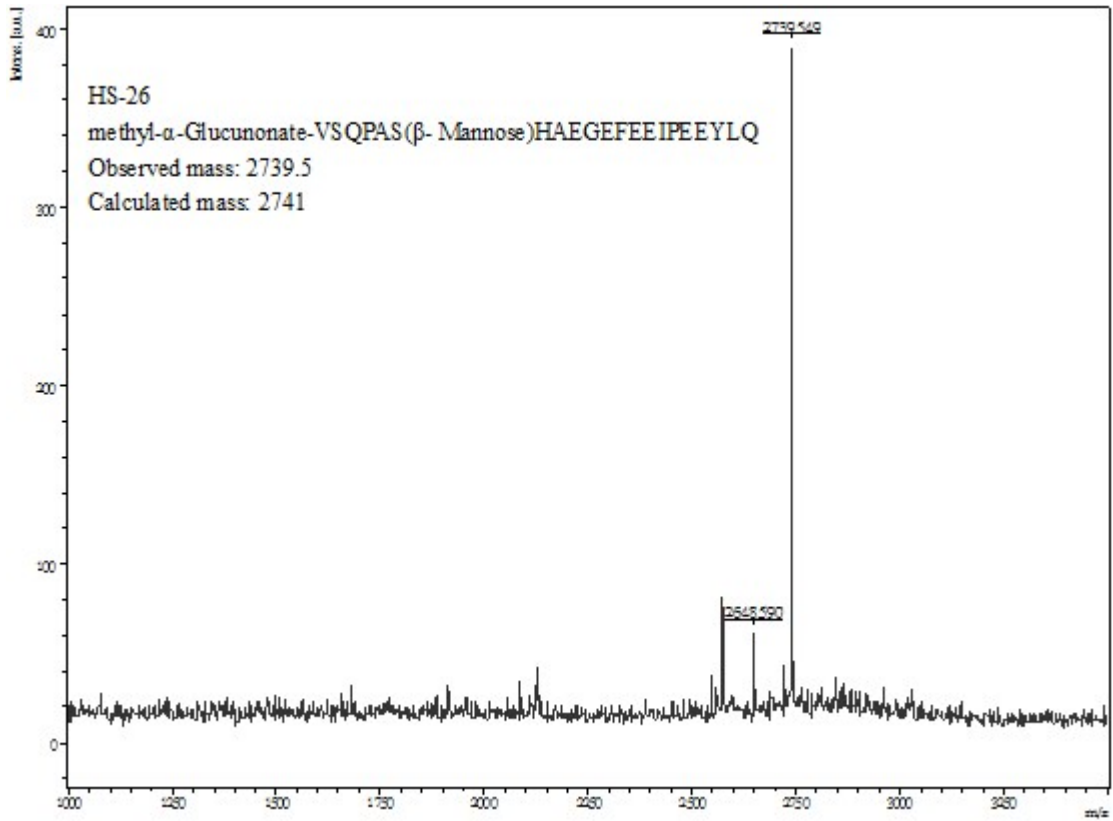
MALDI-TOF mass spectra of HS-23



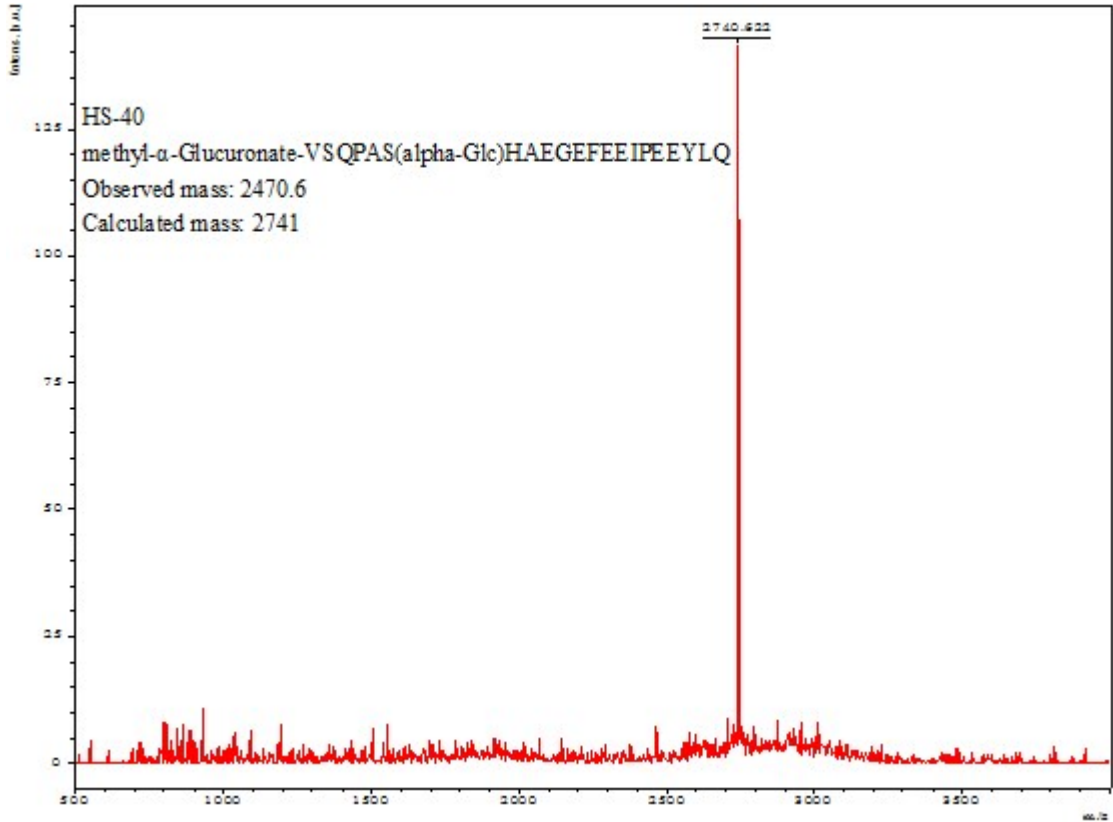
MALDI-TOF mass spectra of HS-24



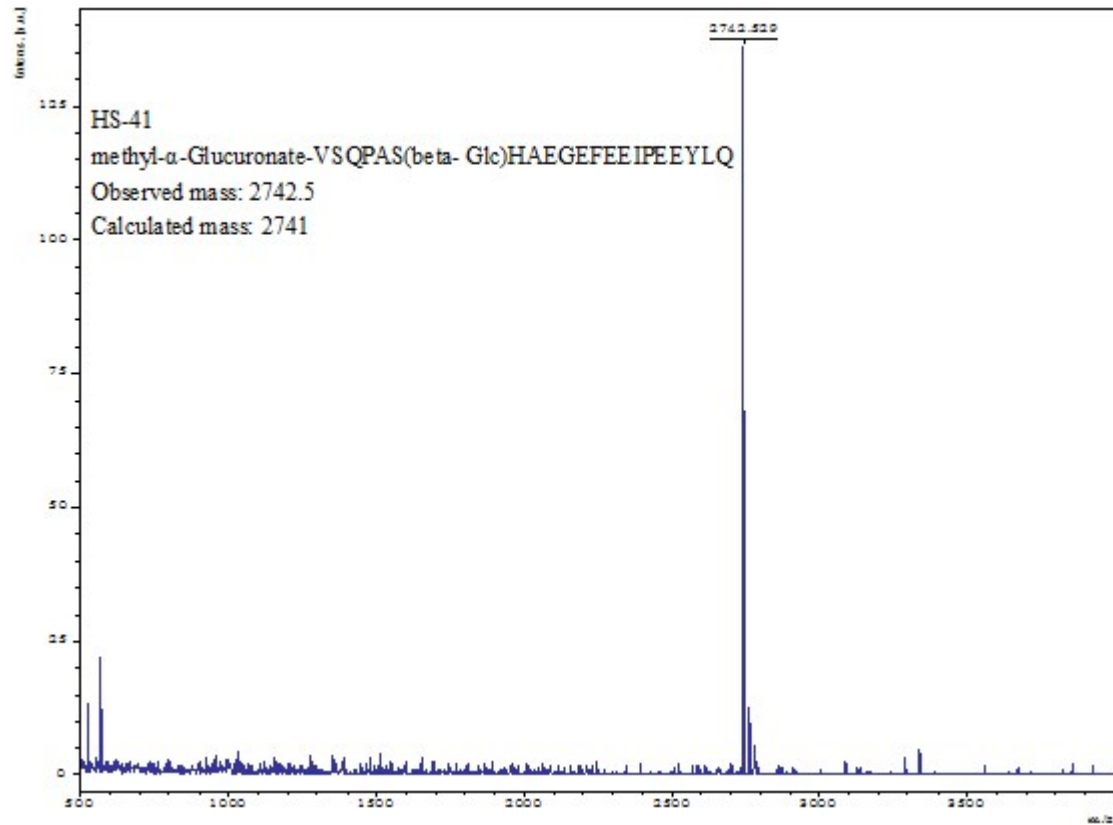
MALDI-TOF mass spectra of HS-25



MALDI-TOF mass spectra of HS-26



MALDI-TOF mass spectra of HS-40



MALDI-TOF mass spectra of HS-41