Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supplementary Information

for

Twisted amide electrophiles enable cyclic peptide sequencing

*Serge Zaretsky,[1] Vishal Rai,[1,] Gerald Gish,[2] Matthew W. Forbes,[1] Michael Kofler,[2] Joy Yu,[1] Tony Pawson,[2] and Andrei K. Yudin[1,]**

[1]Davenport Research Laboratories, Department of Chemistry, University of Toronto, 80 St. George Street, Toronto, Ontario, Canada M5S 3H6

[2]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

Methods

General Information: All solvents including TFE (2,2,2,-trifluoroethanol) and HFIP (1,1,1,3,3,3-hexafluoro-2-isopropanol) were of reagent grade quality.

Nuclear magnetic resonance spectra: ^{1}H and ^{13}C NMR spectra were recorded on Varian Mercury 400 and 500 MHz or Varian 600 MHz spectrometers. ^{1}H NMR spectra were referenced to TMS (0 ppm), CD_3OD (3.30 ppm) and ^{13}C NMR spectra were referenced to CDCl3 (77.2 ppm) and CD_3OD (49.0 ppm). Cyclic peptides aggregate at concentrations higher than 0.5-3 mM. Peak multiplicities are designated by the following abbreviations: s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet; ds, doublet of singlets; dd, doublet of doublets; ddd, doublet of doublet of doublets; bt, broad triplet; td, triplet of doublets; tdd, triplet of doublets of doublets.

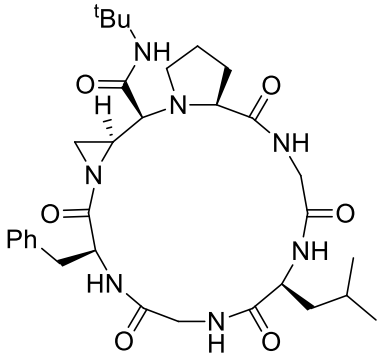
Mass Spectrometry: High-resolution mass spectra were obtained on a VG 70-250S (double focusing) mass spectrometer at 70 eV on a QStar XL (AB Sciex, Concord, ON, Canada) mass spectrometer with electrospray ionization (ESI) source, MS/MS and accurate mass capabilities. Low resolution mass spectra (ESI) were obtained at 60 eV, 70 eV and 100 eV.

RP-HPLC/MS Method A: Low resolution mass spectra (ESI) were collected on an Agilent Technologies 1200 series HPLC paired to a 6130 Mass Spectrometer. Compounds were resolved on an Agilent Poroshell 120 EC-C18, 2.7 μm, 4.6 x 50 mm column at room temperature with a flow of 1 mL/min. The gradient consisted of eluents A (0.1% Formic acid in double distilled water) and B (0.1%formic acid in HPLC-grade acetonitrile). The gradient method started at 5 % of B for the first 0.5 minutes, followed by a linear gradient from 5% to 95% B in 4.0 minutes. The column was then washed with 95 % B for 1.0 minutes and equilibrated at 5 % B for 1.0 minutes.

RP-HPLC/MS Method B: Low resolution mass spectra (ESI) were collected on an Agilent Technologies 1200 series HPLC paired to a 6130 Mass Spectrometer. Compounds were resolved on an Agilent Poroshell 120 EC-C18, 2.7 μm, 4.6 x 50 mm^2 column at room temperature with a flow of 1 mL/min. The gradient consisted of eluents A (0.1% Formic acid in double distilled water) and B (0.1% Formic acid in HPLC-grade acetonitrile). The gradient method started at 5 % of B for the first 0.99 minutes, followed by a linear gradient from 5% to 95% B in 8.0 minutes. The column was then washed with 95 % B for 1.0 minutes and equilibrated at 5 % B for 1.5 minutes.

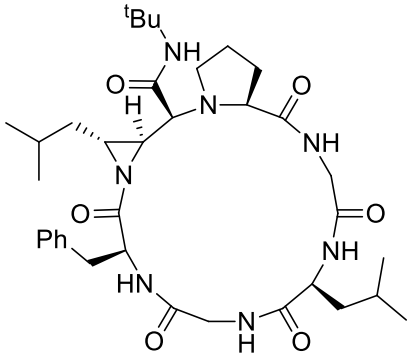
MS^2 Peptide Sequencing*:* Peptide fragmentation spectra were measured using a QStar XL (AB Sciex, Concord, ON, Canada) quadrupole time-of-flight mass spectrometer (Q-TOF-MS) operated in the positive ion mode. Peptide solutions were prepared in a 1:1 mixture of methanol and aqueous 0.1% formic acid to a concentration of ~1-10 μM and were infused at a flow rate of 5 μL min^{-1} to the ESI source. The mass spectrometer is equipped with an HSID atmosphereic pressure interface (IONICS Mass Spectrometry Group Inc., Bolton, ON, Canada). Monoisotopically-resolved precursor ion populations were isolated in the mass-selective

quadrupole (Q1) and fragmented in the collision cell (Q2) having an elevated pressure of N_2 collision gas. Product ion spectra were recorded for several minutes at a variety of collision voltages to optimize fragment ion intensities and signal-to-noise ratio. The peptide fragmentation spectra were analyzed using BioAnalyst v. 1.1 (AB Sciex).



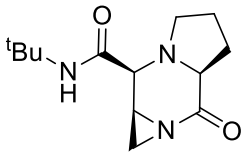
(4S,10S,15aS,20S,20aS)-4-benzyl-N-(tert-butyl)-10-isobutyl-3,6,9,12,15-pentaoxoicosahydroazirino[1,2-a]pyrrolo[1,2-d][1,4,7,10,13,16]hexaazacyclooctadecine-20-carboxamide (2a)

Reported in *J. Am. Chem. Soc.*, 2010, **132**, 2889 and *Chem. Eur. J.*, 2013, **19**, 17668.



(1R,4S,10S,15aS,20S,20aS)-4-benzyl-N-(tert-butyl)-1,10-diisobutyl-3,6,9,12,15-pentaoxoicosahydroazirino[1,2-a]pyrrolo[1,2-d][1,4,7,10,13,16]hexaazacyclooctadecine-20-carboxamide (2b)

Reported in *J. Am. Chem. Soc.*, 2010, **132**, 2889 and *Chem. Eur. J.*, 2013, **19**, 17668.

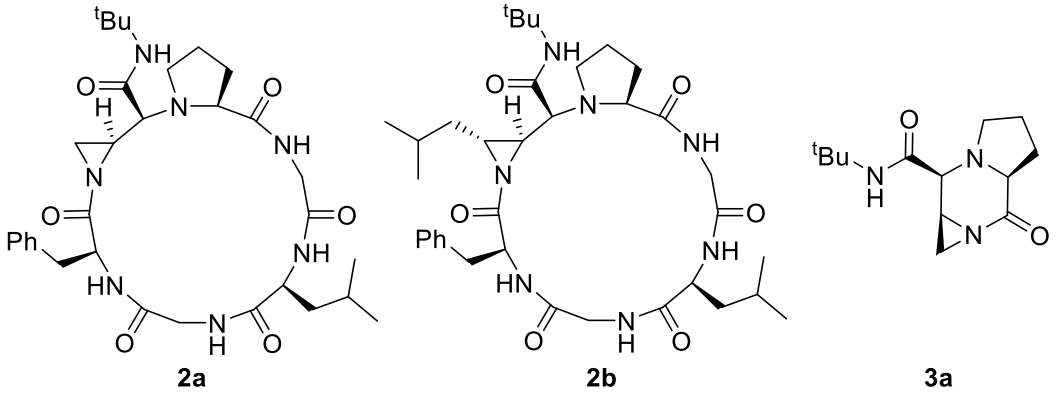


(3aS,8S,8aS)-N-(tert-butyl)-3-oxooctahydroazirino[1,2-a]pyrrolo[1,2-d]pyrazine-8-carboxamide (3a)

Reported in L. Belding, S. Zaretsky, B. H. Rotstein, A. K. Yudin, and T. Dudding, *J. Org. Chem.*, 2014, **79**, 9465–9471.

Investigation of reactivity against biological nucleophiles

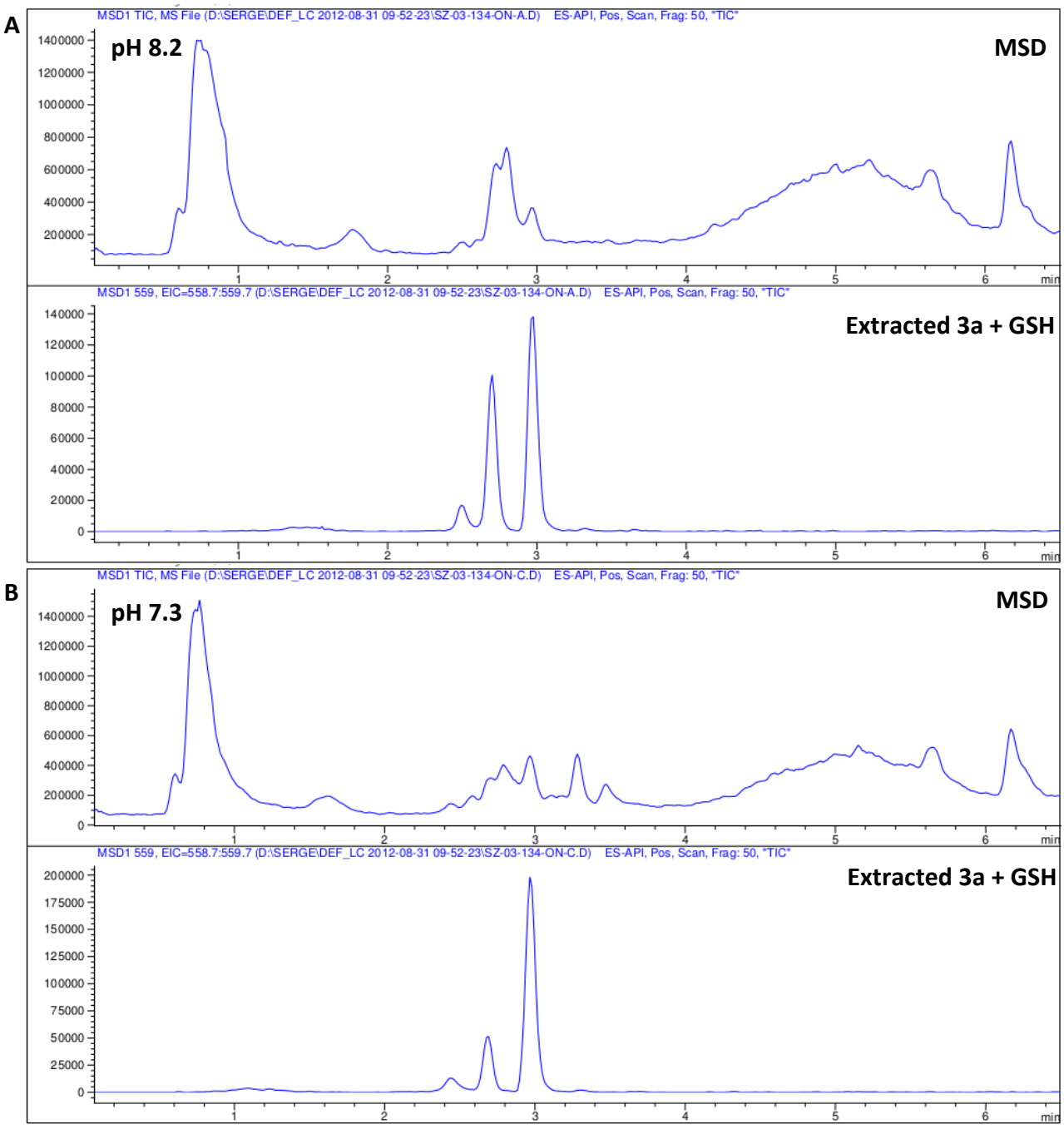
Peptide cycles produced in the disrupted Ugi macrocyclization contain electrophilic aziridine amides. For the synthesis/sequencing methodology described in this paper to find application, we needed to evaluate whether or not aziridine-containing cycles would react with biological nucleophiles by either S_N2 or transamidation pathway. For our study, we chose representative molecules **2a**/**2b**, and **3a** (**Supplementary Figure 1**) that were synthesized using the previously reported conditions. The reactivity of these aziridine amide-containing cycles towards biological nucleophiles, found in L-glutathione (GSH) and the catalytic domain of the protein tyrosine kinase Fes, was investigated.



Supplementary Figure 1. Representative cyclic aziridine-containing molecules for testing of irreversible labeling with biological nucleophiles.

A glutathione study

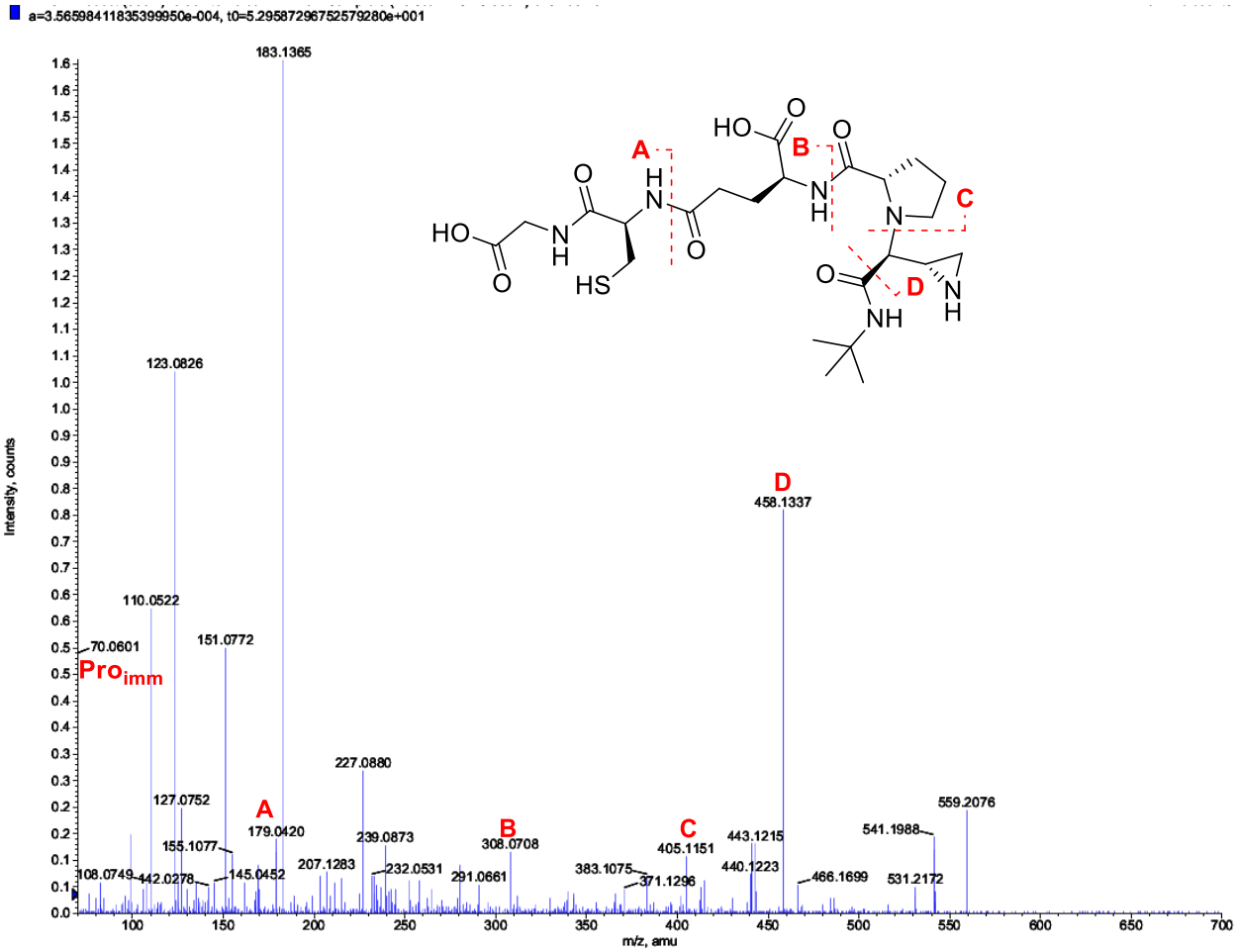
Two reaction vials were charged with compound **3a**. 100 mM GSH adjusted to pH 8.2 with ammonium bicarbonate and 100 mM GSH adjusted to pH 7.3 with additional HCl (1.0 M) was added to the respective vials. The resulting mixtures were incubated for 18 hours at room temperature. Reaction products were analyzed by mass spectrometry directly using RP-HPLC/MS Method A and B. A mixture of GSH adducts was observed at pH 8.2, but at pH 7.3 the reaction was more selective for the later eluting product (**Supplementary Figure 2**).



Supplementary Figure 2. HPLC/MS spectra of GSH-**3a** adducts at pH 8.2 (a) and 7.3 (b). At pH 7.3 there was greater selectivity for one type of GSH adduct (MS (ESI) [M+H]+ calcd. 559.3, found 559.2. HPLC/MS Method A retention time of 2.969 vs. 2.698).

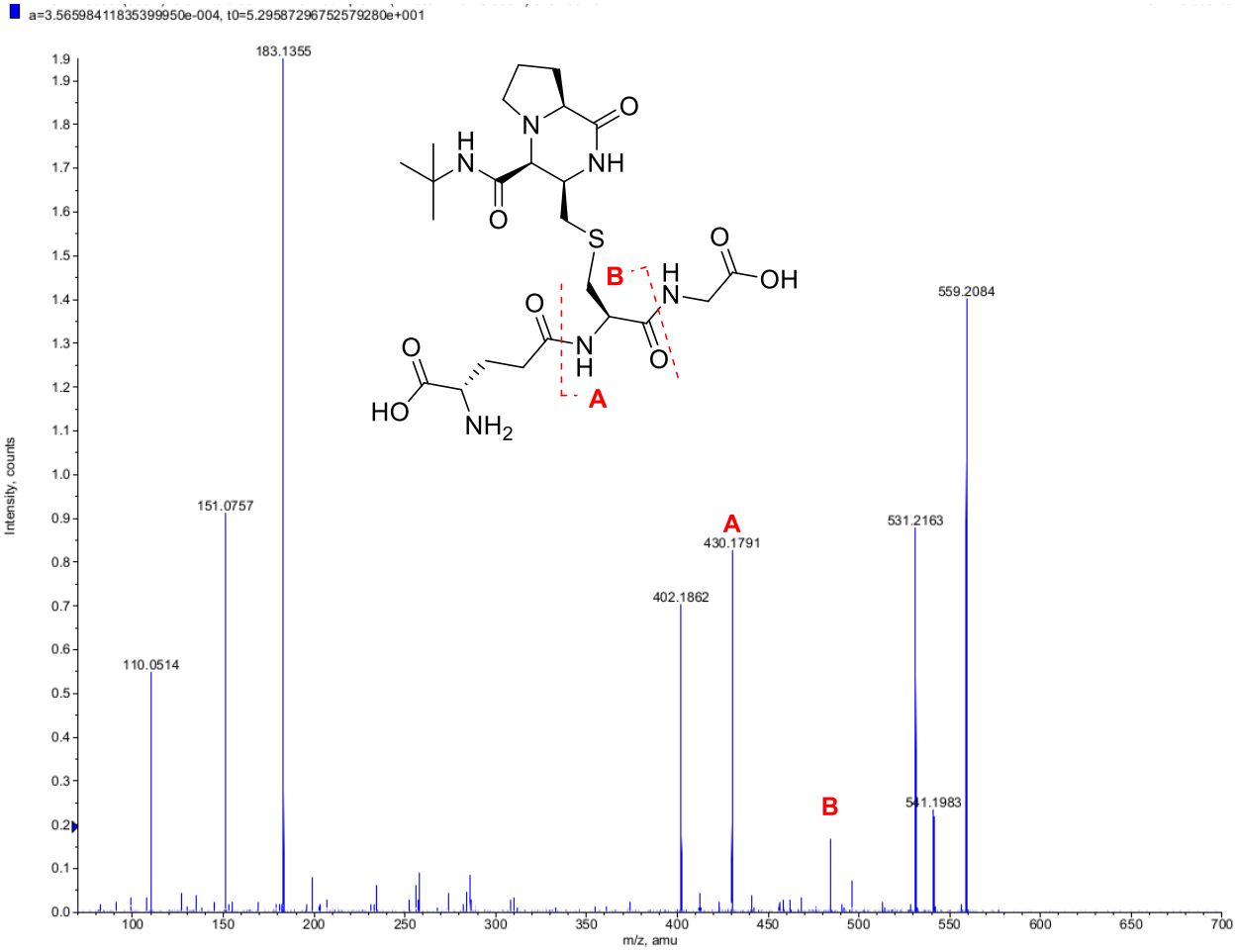
The isobaric peaks at pH 8.2 were individually collected from an analytical HPLC/MS Method A run and subjected to fragmentation by CID/MS^2 to deduce the type of nucleophilic

transformation. The later eluting peak fragmented in a similar manner to conventional linear peptide fragmentation. Two *y*-ions of GSH were observed as well as the transamidated adduct of GSH and **3a** (**Supplementary Figure 3**).

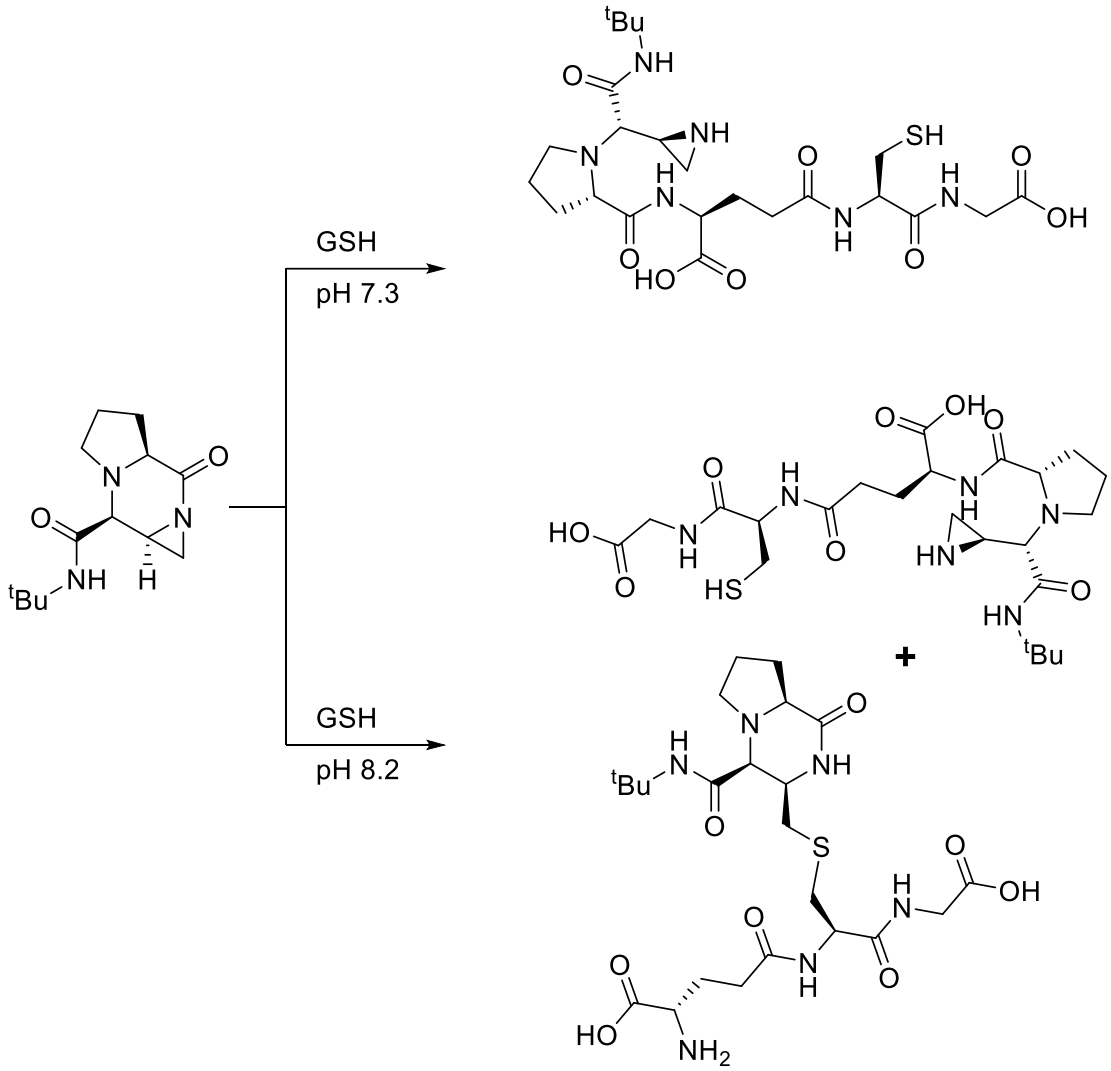


Supplementary Figure 3. CID/MS^2 spectrum of GSH-**3a** adduct at pH 8.2 and 7.3.

The pH 8.2 selective peak (earlier elution time) fragmented very differently by CID and a series of very prominent ions was observed (**Supplementary Figure 4**). The absence of fragments within the core of **3a**, and only fragmentation on either side of the Cys residue of GSH, led to the conclusion that the GSH-**3a** adduct with the retention time of 2.698 (HPLC/MS Method A) was the S_N2 product of sulfhydryl attack on the aziridine (**Supplementary Figure 5**).



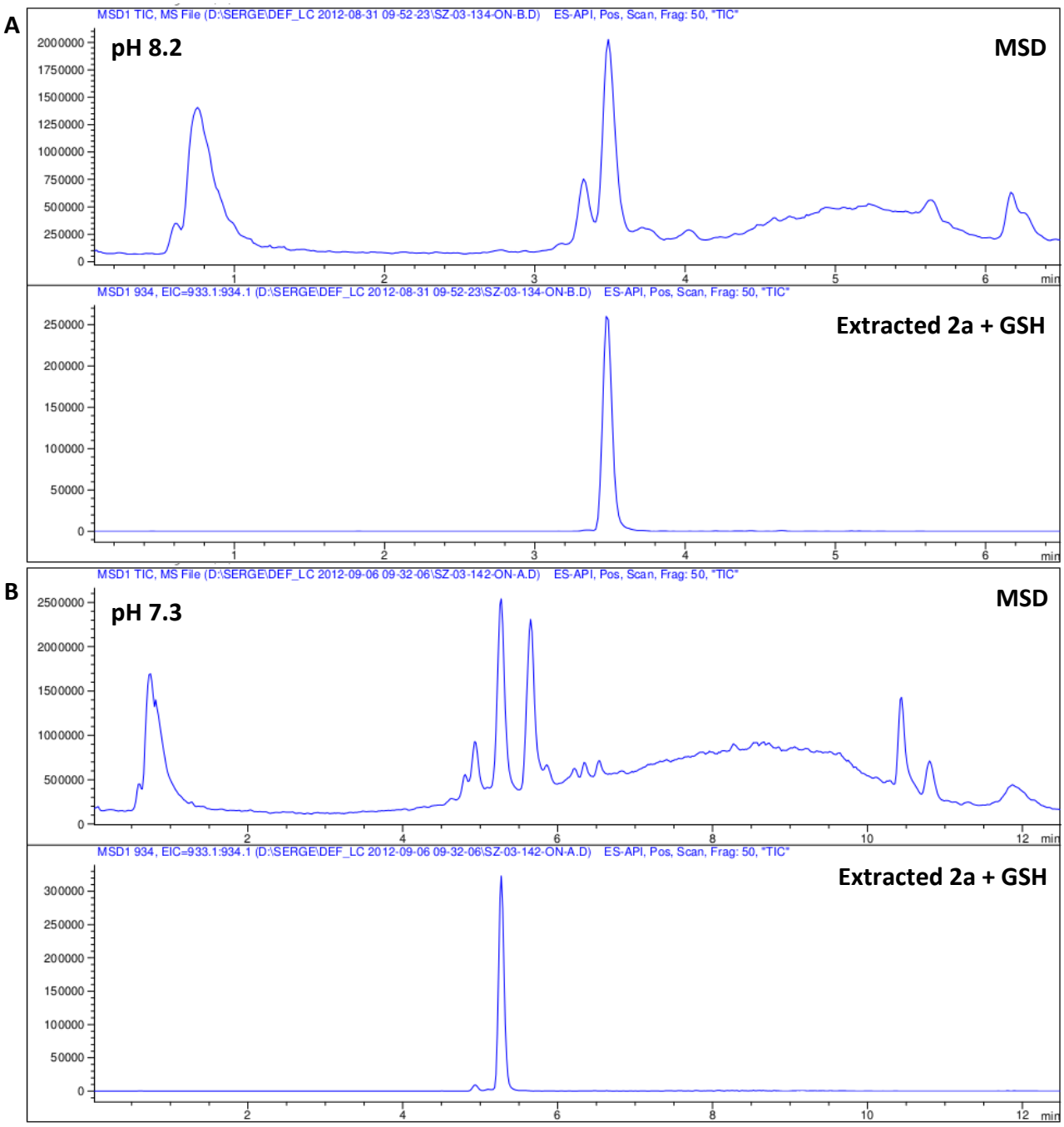
Supplementary Figure 4. CID/MS^2 spectrum of pH 8.2 selective GSH-**3a** adduct.



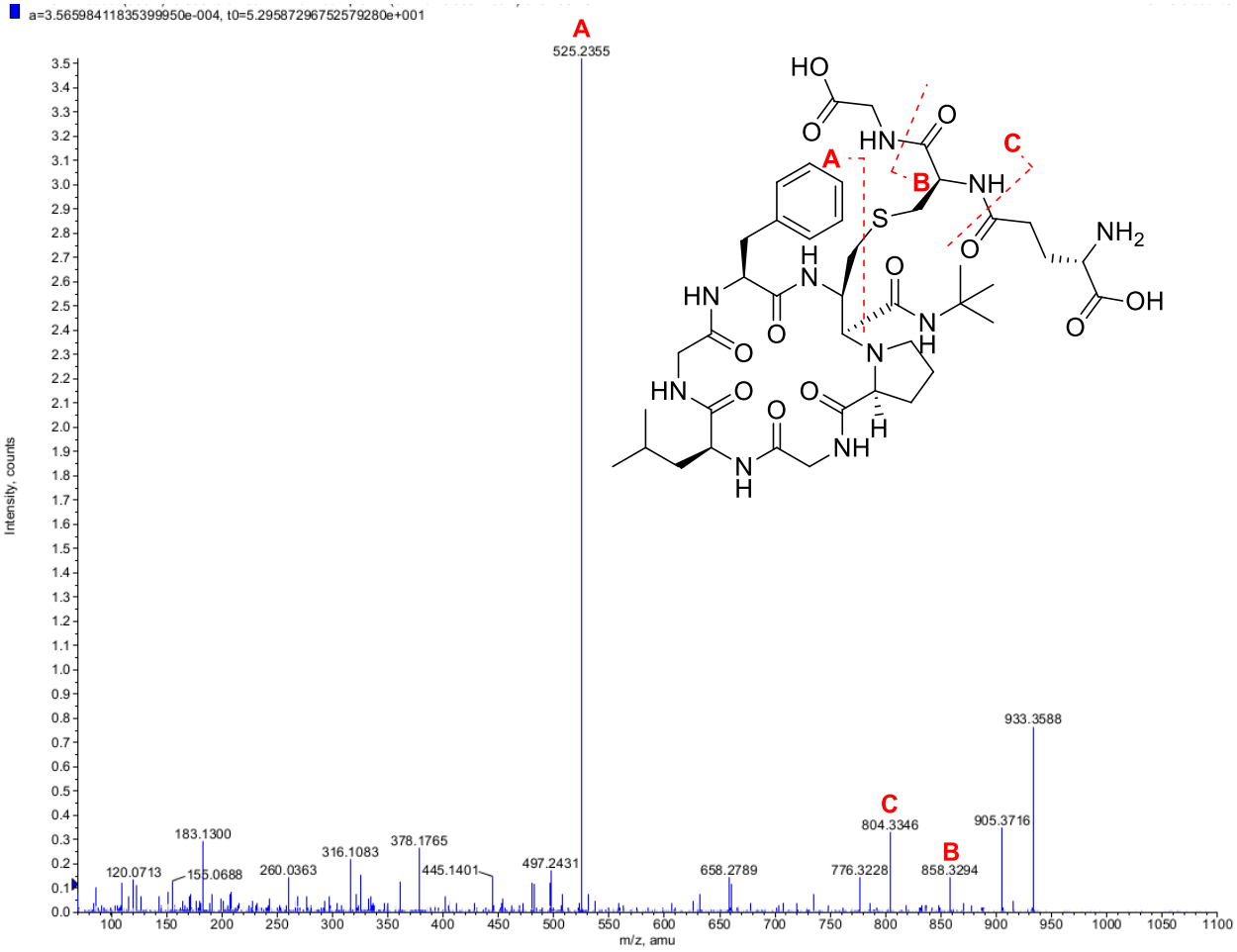
Supplementary Figure 5. Basic conditions lead to S_N2 reactivity while at neutral pH, the transamidation pathway is predominant.

Under similar conditions to the pH dependent testing of **3a**, **2a** and **2b** were screened for their reactivity with GSH. Reaction products were once again analyzed by mass spectrometry directly using RP-HPLC/MS Method A and B. Only one GSH-**2a** adduct was observed at both pH 7.3 and 8.2 (**Supplementary Figure 6**). When exposed to CID/MS^2, GSH-**2a** fragmented only along GSH and no core linear peptide fragments were observed (**Supplementary Figure 7**), thereby indicating S_N2 reactivity (**Supplementary Figure 8**). Under prolonged exposure at pH 8.2, hydrolysis of **2a** was observed as well as a GSH adduct of the

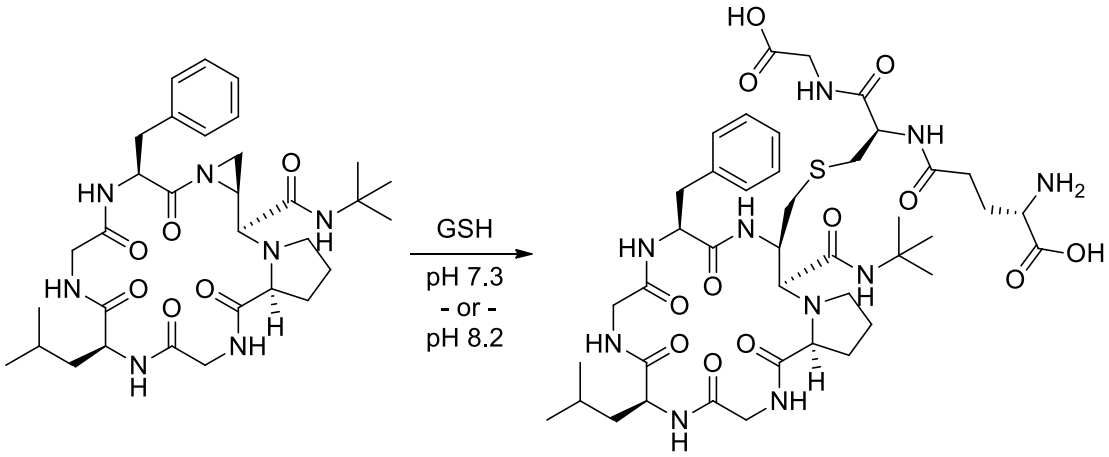
hydrolyzed peptide. The sterically encumbered isobutyl derivative **2b**, failed to react with GSH under either pH conditions (**Supplementary Figures 9-10**).



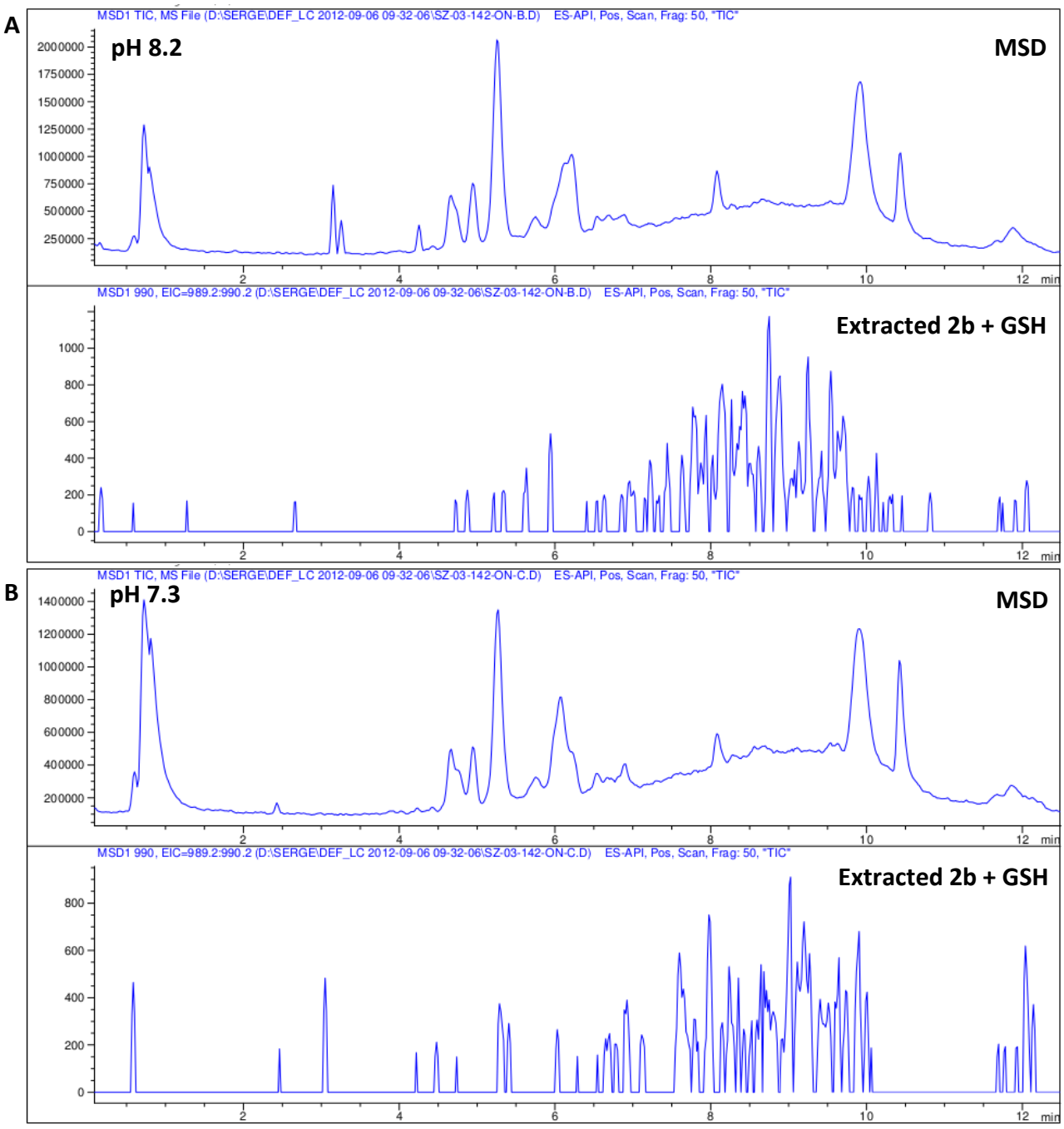
Supplementary Figure 6. HPLC/MS spectra of GSH-**2a** adduct at pH 8.2 (a) and 7.3 (b). Only one main GSH-**2a** peak was observed (MS (ESI) [M+H]+ calcd. 933.5, found 933.4. RP-HPLC Method A retention time 3.481 min, RP-HPLC Method B retention time 5.272 min).



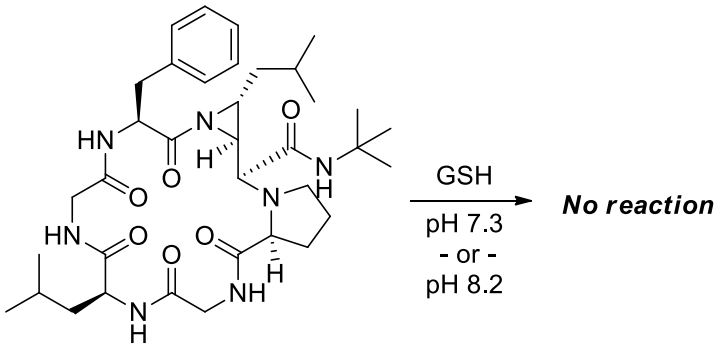
Supplementary Figure 7. The GSH-**2a** adduct fragmented only along GSH and no peptide backbone fragments were observed.



Supplementary Figure 8. 2a reacts with GSH only by an S_N2 pathway.



Supplementary Figure 9. HPLC/MS spectra did not indicate formation of GSH-**2b** adducts at pH 8.2 (a) or 7.3 (b).



Supplementary Figure 10. Sterically encumbered acyl aziridine-containing macrocycle **2b**, fails to react with GSH.

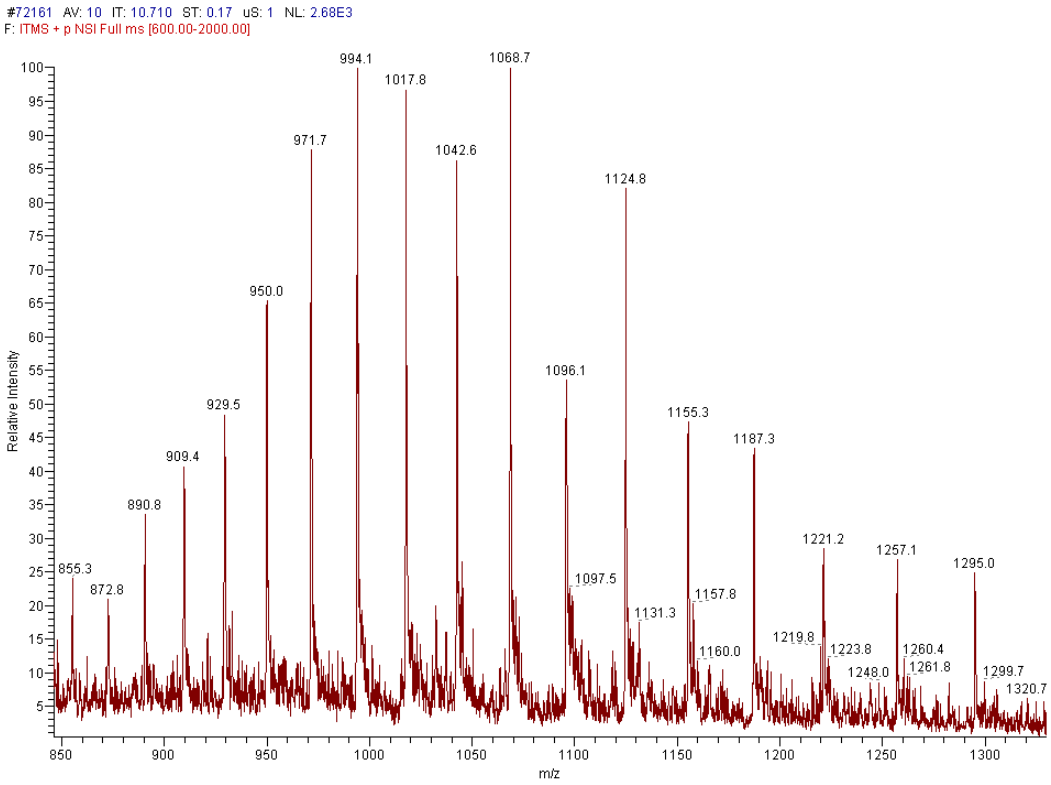
A Fes SH2-kinase protein study

Protein expression and purification: The FesA expression construct was transformed into phage resistant *E. coli* BL21(DE3)-R3 co-transformed with an expression vector encoding *Yersinia* phosphatase YopH.[1] Cells were grown in Terrific Broth medium containing 50 µg/ml kanamycin and 50 µg/ml streptomycin at 37 °C to an optical density of about 3.5 (OD600) before the temperature was decreased to 18 °C for a 12h induction of protein expression through addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside. Cells were re-suspended in lysis buffer (100 mM NaH_2PO4, pH 8.0 at 25 °C, 1 M NaCl, 5% glycerol, 0.5 mM tris(2-carboxyethyl) phosphine (TCEP), 10 mM imidazole) in the presence of a protease inhibitor cocktail tablet mix (Complete, EDTA-free Protease Inhibitor Cocktail, Roche Diagnostics Ltd.) and lysed using an EmulsiFlex-C5 high pressure homogenizer (Avestin) at 4 °C. The lysate was incubated with 0.15% polyethylenimine (PEI) at pH 7.5 for 30 min at 4 °C in order to remove residual DNA, and cleared by centrifugation and filtration through a 0.2 µm serum Acrodisc filter. Protein purification was performed using sequential FPLC His-trap (Pharmacia) and size exclusion chromatography (SEC) (Superdex 75 16/60 HiLoad, GE/Amersham Biosciences) at 7 °C. The His-trap column was washed with 50 ml binding buffer (50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole pH 8.0, 5% glycerol, 0.5 mM TCEP), 100 ml wash buffer (50 mM NaH_2PO_4, 500 mM NaCl, 20 mM imidazole, pH 8.0, 5% glycerol. 0.5 mM TCEP), and recombinant Fes SH2-kinase was eluted using elution buffer (50 mM NaH_2PO_4, 500 mM NaCl, 300 mM imidazole pH

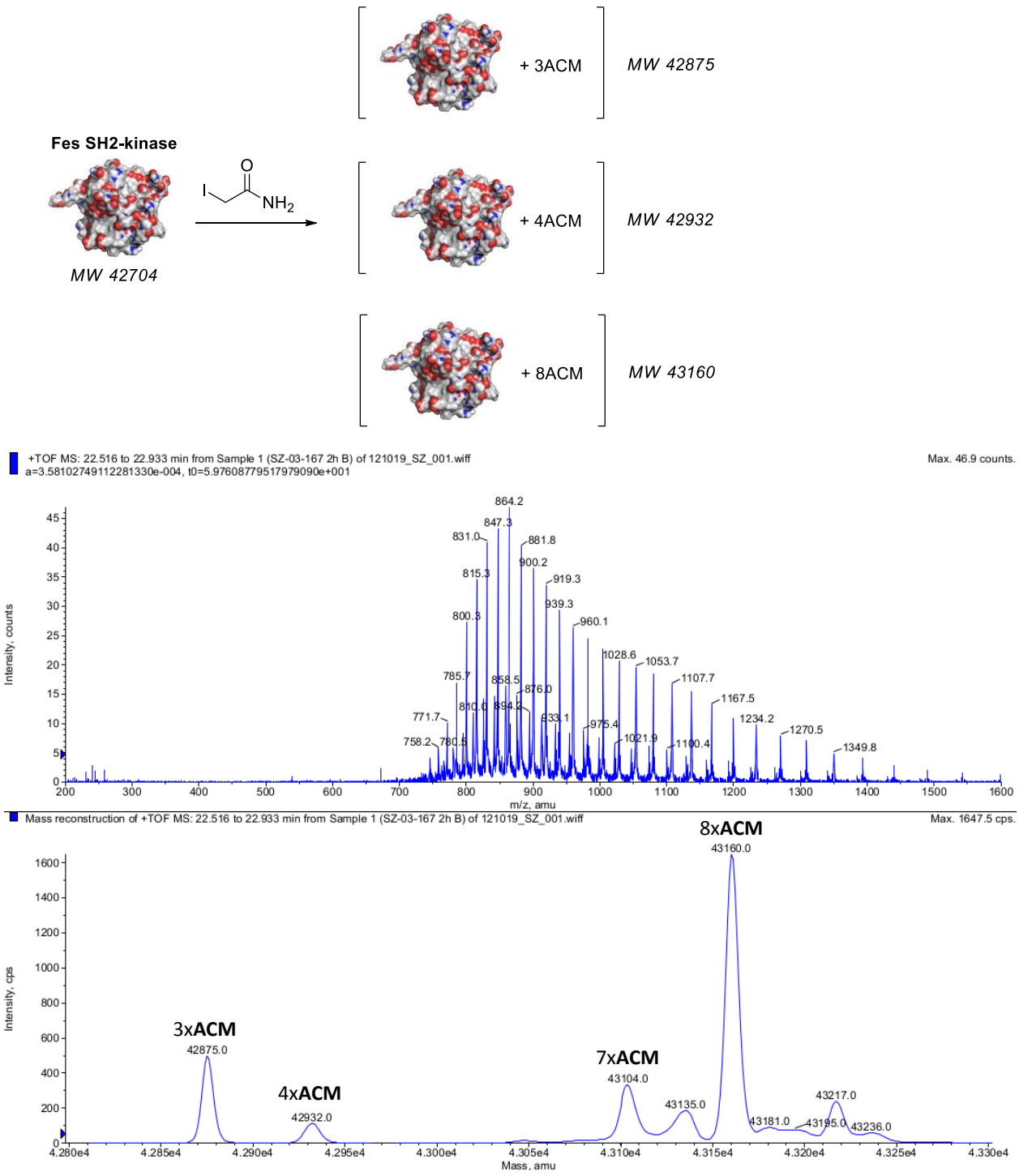
8.0, 5% glycerol, 0.5 mM TCEP). The peak fraction was further purified by SEC eluting in 10 mM HEPES, pH 7.4, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP. The protein was treated for 12 h at 4 ^{o}C with TEV protease to remove the hexa-histidine tag. The cleaved tag and uncleaved protein were removed using a nickel nitrilotriacetic acid agarose column. Purity was monitored by SDS-polyacrylamide gel electrophoresis and final samples were concentrated to 5 mg/ml (see **Supplementary Figure 11** for MS analysis).

Fes SH2-kinase (5 µL, 25 µM, resuspended in 25 mM Tris HCl (pH 7.5), 150 mM NaCl) was added to a 0.5 mL microcentrifuge tube charged with **3a** (0.5 µL, 0.1 mM in DMSO). The resulting mixture was incubated at room temperature. Reaction products were analyzed by mass spectrometry directly using an Agilent Poroshell 120 EC-C8, 2.7 µm, 2.1 x 150 mm column at room temperature with a flow of 0.200 mL/min. on a QStarXL ToF spectrometer and analyzed with Analyst. Protein reconstruction with a scan from 30 to 60 kDa was used to reconstruct protein masses from ions which lay mostly in the 700-1000 M/z range. At 3 hours, an aliquot of 1 µL was analyzed and then iodoacetamide (0.5 µL, 0.1 mM in DMSO) was added to the remaining reaction and left to incubate for an additional 3 hours at room temperature, before subsequent analysis.

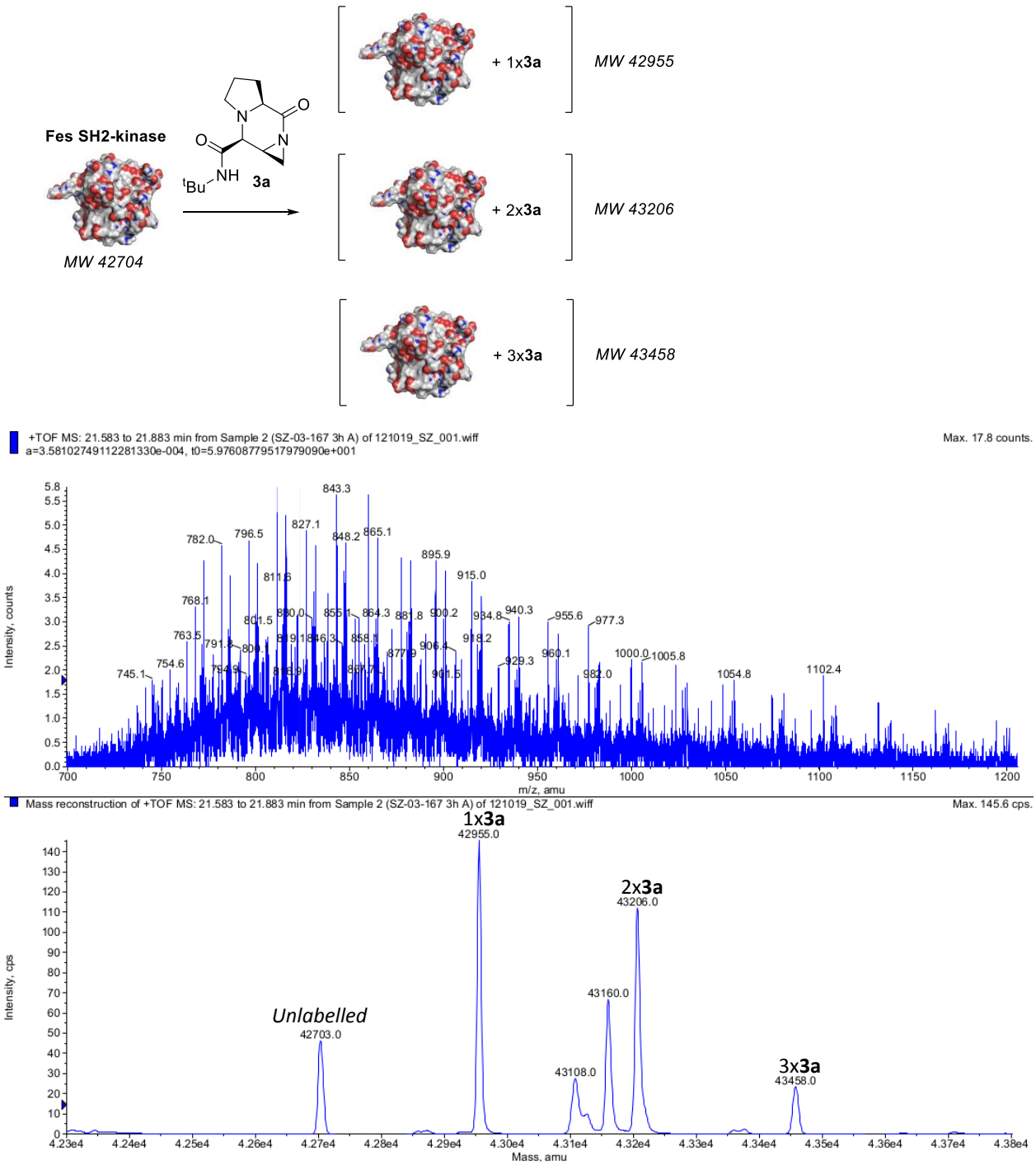
In a control reaction, Fes SH2-kinase (5 µL, 25 µM, resuspended in 25 mM Tris HCl (pH 7.5), 150 mM NaCl) was only exposed to iodoacetamide (0.5 µL, 0.1 mM in DMSO) (**Supplementary Figure 12**) and up to 8 adducts of acetamide (ACM) groups were observed after 2 hours of incubation at room temperature. Exposure to **3a** generated up to 3 adducts after 3 hours of exposure to Fes SH2-kinase (**Supplementary Figure 13**). Upon treatment of Fes SH2-kinase exposed to **3a**, with iodoacetamide the protein was further alkylated with ACM groups. As in the control, the iodoacetamide alkylation pattern was preserved, with primarily 3 and 8 ACM gains (**Supplementary Figure 14**).



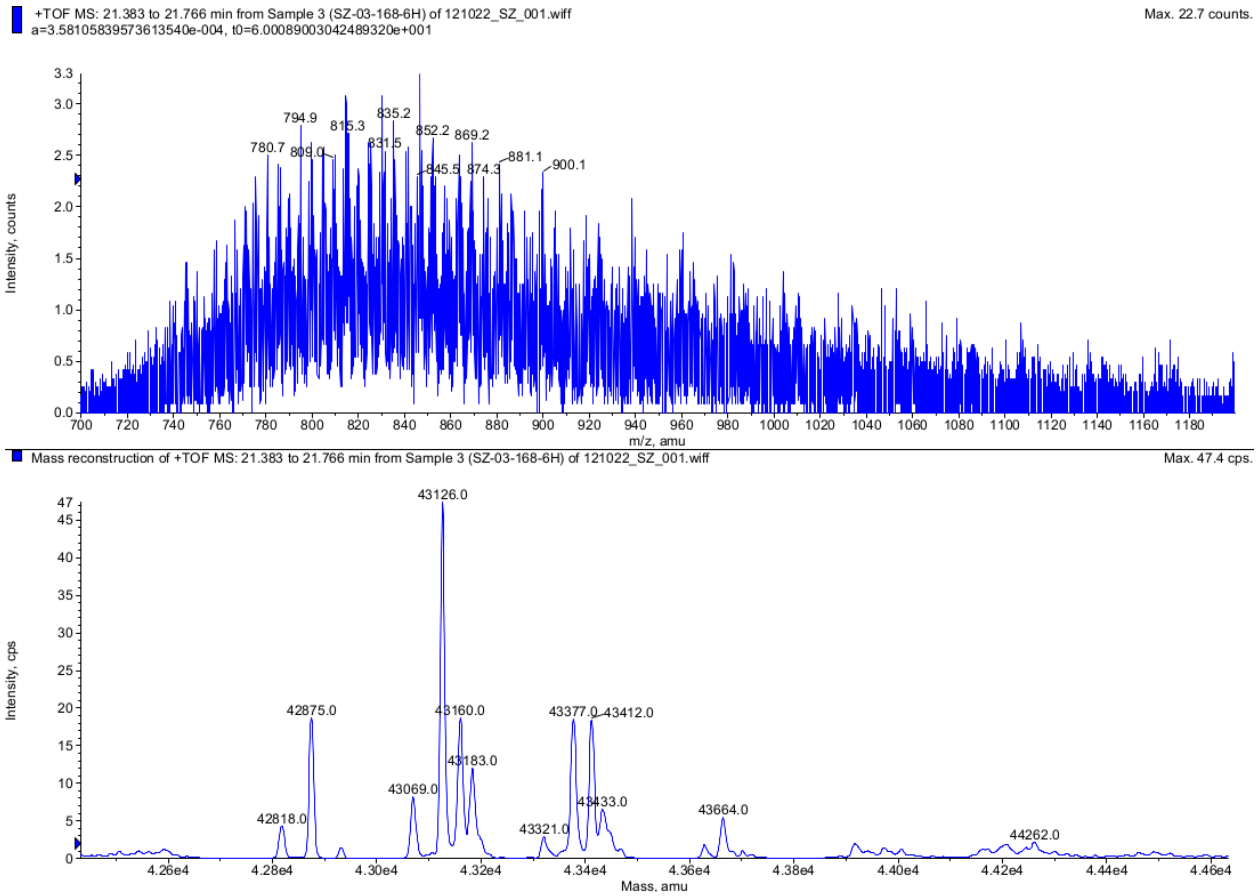
Supplementary Figure 11. The ion spectrum for the unmodified Fes SH2-kinase (molecular weight ≈ 42,705) captured the protein in the +33 to +50 charged states, 1295.0 to 855.3 M/z respectively.



Supplementary Figure 12. The positive control labelling of Fes SH2-kinase with iodoacetamide showed significant alkylation of the Cys residues (primarily observed as the 3x, 4x, 7x, and 8x alkylated species).

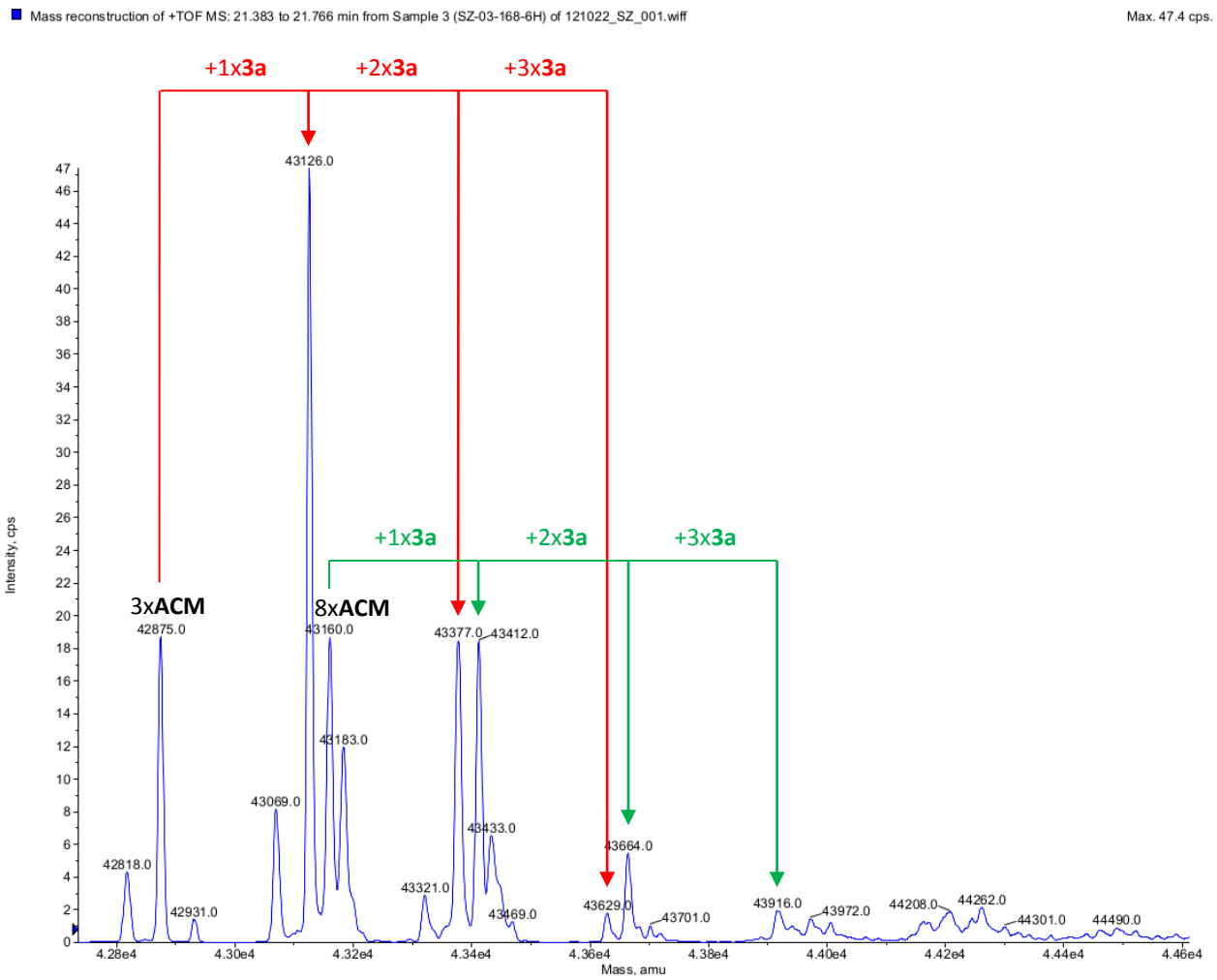


Supplementary Figure 13. Fes SH2-kinase exposed to **3a** for 3 hours was primarily observed as the unlabelled protein and 1x, 2x, and 3x adducts of **3a**.

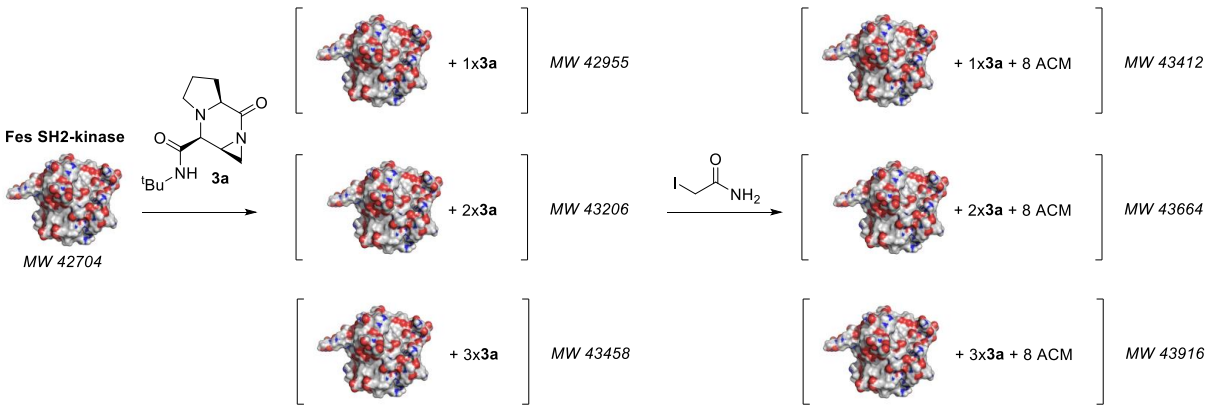


Supplementary Figure 14. Per-alkylation with iodoacetamide of Fes SH2-kinase that had been previously exposed to **3a**.

Further analysis of the iodoacetamide and **3a** exposed Fes SH2-kinase, indicated that the most prominent protein ions corresponded to gains of 3 and 8 ACM groups (**Supplementary Figure 15**). The extent of alkylation with **3a** did not affect the number of ACM groups gains. For example, 43412 corresponds to one molecule of **3a** and 8 ACM, 43664 corresponds to two molecules of **3a** and 8 ACM (**Supplementary Figure 16**). Thus, we concluded that nucleophilic reactivity of the protein with **3a**, does not prevent alkylation with iodoacetamide and thus is not mediated by cysteine residues.

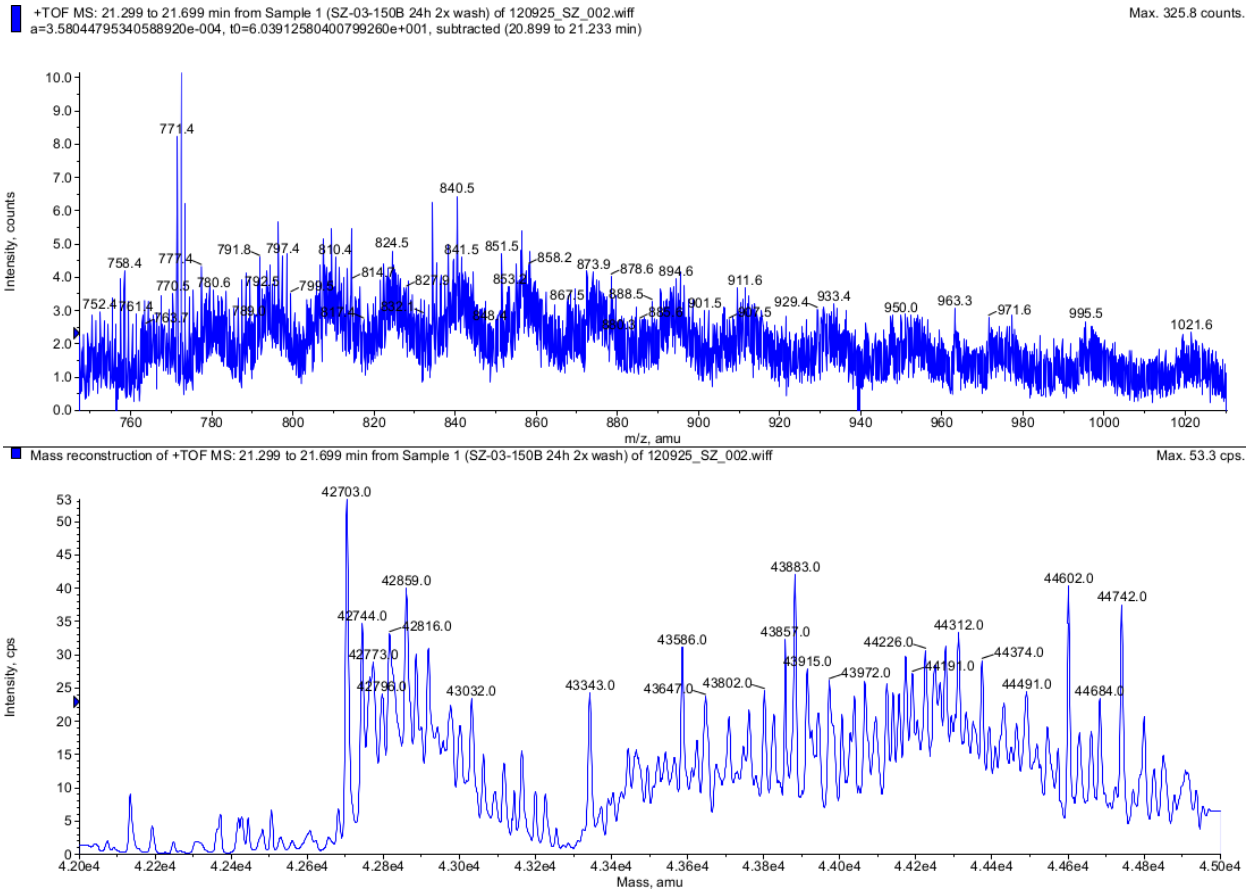


Supplementary Figure 15. Fes SH2-kinase is alkylated with iodoacetamide in an identical manner, regardless of the number of **3a** adducts.



Supplementary Figure 16. Nucleophilic reactivity of Fes SH2-kinase does not prevent alkylation with iodoacetamide.

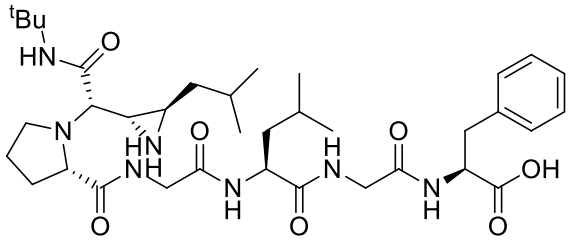
Knowing that larger ring sizes shut down electrophilicity at the aziridine amide carbonyl by increasing the n-to-π* interaction (as per previous GSH adduct results) we next attempted to alkylate Fes with **2a**. Fes SH2-kinase was exposed to **2a** dissolved in DMSO in an identical manner to the alkylation experiments with **3a**. In contrast to the reactivity seen with **3a**, there was no reaction between Fes SH2-kinase and the cyclic PGLGF peptides **2a** (**Supplementary Figure 17**), even after incubation for 24 hours. Lack of attack at the aziridine implies that the cysteines found on Fes SH2-kinase are not nucleophilic enough or there is poor contact made between the cyclized peptide and protein. The simple change to an 18-membered ring shuts down carbonyl electrophilicity completely. In this case, ring size alone is able to reduce the electrophilicity of the aziridine amide.



Supplementary Figure 17. Fes SH2-kinase was not covalently labelled by **2a** after 24 hours of exposure

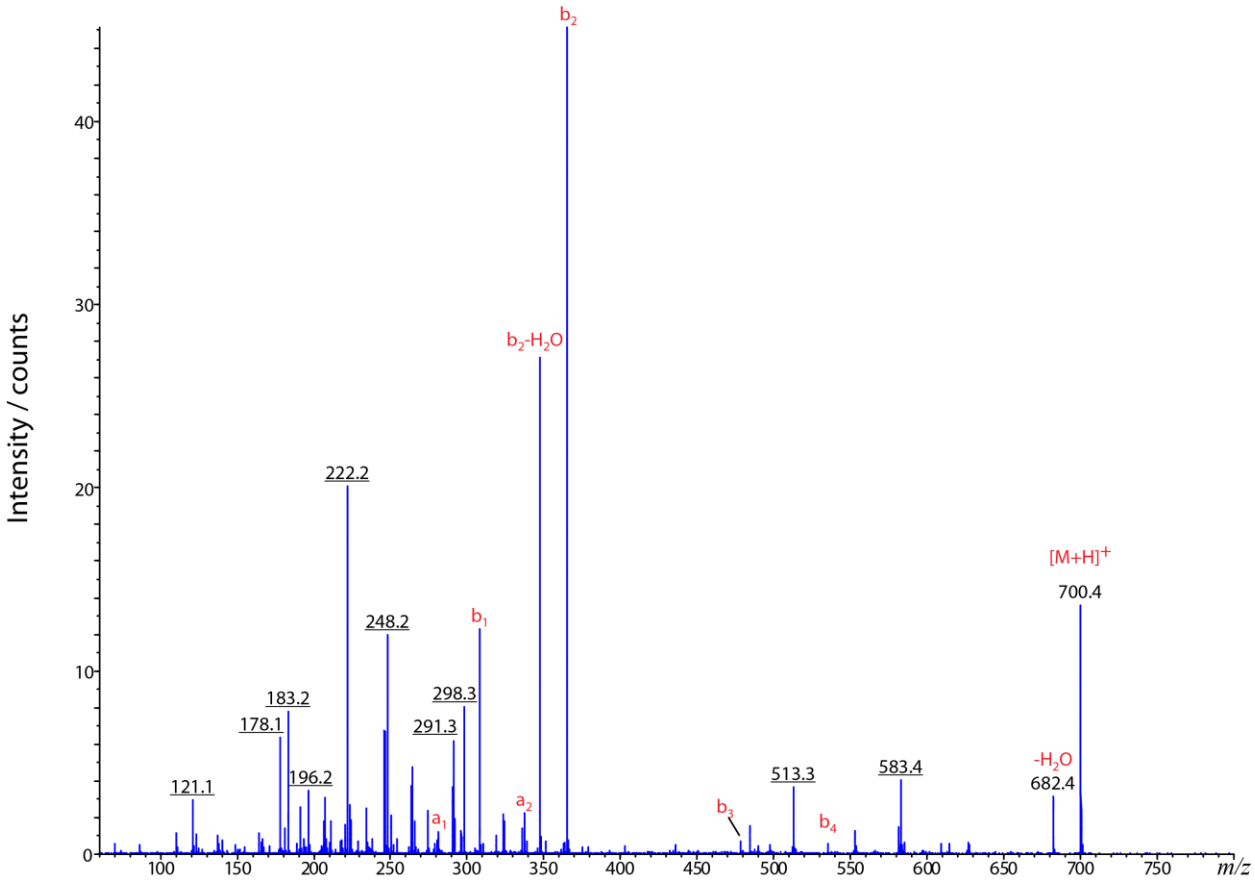
Site-selective hydrolysis and peptide sequencing

To investigate the site-selective hydrolysis with hard nucleophiles, **2b** was synthesized as per literature procedures.[2] **2b** was then linearized under basic conditions.



(6S,12S)-12-benzyl-1-((S)-1-((S)-2-(tert-butylamino)-1-((2S,3R)-3-isobutylaziridin-2-yl)-2-oxoethyl)pyrrolidin-2-yl)-6-isobutyl-1,4,7,10-tetraoxo-2,5,8,11-tetraazatridecan-13-oic acid (4)

To **2b**, (0.5 mg) was added NaOH (50 µL, 1.5 M) and ddH_2O (50 µL). Following stirring for 90 minutes, aliquots of the mixture were analyzed by HPLC-MS and CID/MS^2. ESI MS $[M+H]^+$ expected: 700.4, experimental: 700.4. HRMS $[M+H]^+$ expected: 700.4398, experimental: 700.4380. RP-HPLC Method B retention time 6.376 min.



Supplementary Figure 18. CID/MS^2 spectrum of **4**.

4 fragmented readily by CID/MS^2 (**Supplementary Figure 18**). Masses were determined to four decimal places with an external calibration. Underlined fragment ions were determined to be non-sequence ions. The underlined ions for **4** could be related to the partial fragmentation of the linker as well as the b_2 ion series. These ions could find further utility as fingerprint ions for unambiguous assignment of the b_1 and b_2 ions.

Supplementary Table 1. Fragment ion list for **4**.

Centroid mass (amu)	Intensity (counts)
70.0652	0.8
110.0588	1.6
121.0776	4.3
123.0925	1.5
137.0741	1.5

140.1071	1.1
164.1478	1.8
166.0958	1.2
178.0979	9.3
181.1713	2.1
183.1513	11.5
189.1392	0.9
191.1549	3.8
193.1396	1.2
196.1691	5.0
206.1664	2.8
207.1500	4.5
208.1388	1.2
210.1601	0.9
211.1805	2.6
217.1331	1.0
218.1612	1.1
220.1793	2.3
222.1609	29.2
223.1081	3.8
224.1767	2.6
229.1914	1.0
234.1604	3.8
235.1456	0.8
238.1541	1.2
246.1617	9.9
248.1769	17.2
250.1560	3.2
254.1870	1.2
263.2100	5.7
264.1730	6.8
266.1515	2.7
274.1577	3.5
279.2052	0.8
280.2388	1.2
281.2246	1.6
290.2231	5.4
291.1823	9.0
292.1655	2.8
296.2335	2.0
298.2506	11.8
308.2349	17.4

309.2102	0.8
310.2532	0.8
319.2422	1.3
323.2463	3.2
324.2544	2.7
336.1941	2.1
337.2595	3.2
339.2041	1.0
347.2458	39.5
348.2409	1.5
351.2378	0.9
363.2389	1.0
365.2567	65.4
366.2561	1.2
478.3468	1.0
484.2564	2.4
497.2728	0.8
513.2813	5.3
535.3551	0.8
553.3705	2.0
581.3383	2.1
583.3573	5.5
585.3394	1.0
614.3589	1.0
627.3482	0.9
682.4275	4.3
700.4381	19.1

Peptide sequencing was performed in tag mode with a mass tolerance of 0.01 Da. The screened amino acid library was composed of the 20 most common natural amino acids and *P added as a modified Pro residue with the cyclization linker. The only other tagged full sequence was that of PGAVF, an isobaric match for **4** (**Supplementary Table 2**).

Supplementary Table 2. Tagged MS^2 sequences of **4**.

Tag	Score
*PQVF(-0.0018)	82.02
F(534.3511)	81.61
FK(406.2562)	80.65

(325.2385)E(227.1384)	78.98
(364.2493)LGF(-0.0037)	78.80
(534.3473)F(0.0038)	77.88
***PGLGF(-0.0018)**	77.77
(325.2385)EQV(0.0114)	77.76
(364.2493)L(204.0862)	77.26
(364.2493)AVF(-0.0037)	77.03
(477.3392)GF(-0.0096)	76.81
*PGL(204.0880)	76.56
(435.2749)VF(0.0078)	76.54
(454.2763)QV(0.0163)	76.47
*PGAVF(-0.0018)	74.07
(364.2493)A(246.1331)	73.07
*PG(317.1721)	71.65
*PGA(246.1350)	68.22
(364.2493)AFV(-0.0037)	68.22
(435.2749)FV(0.0078)	66.36

The calculated error for informative fragment ions of **4** was less than 10 ppm for all but the b_3 ion, which had an error of 17 ppm (**Supplementary Table 3**).

Supplementary Table 3. Calculated error in most significant fragment ions for **4**

Ion	Calculated mass (amu)	Experimental mass (amu)	Mass difference (amu)	Error (ppm)
[**4**+H]$^+$	700.4398	700.4380	0.0018	2.6
a_1	280.2383	280.2388	-0.0005	1.8
a_2	337.2598	337.2593	0.0005	1.5
b_1	308.2333	308.2350	-0.0017	5.5
b_2	365.2547	365.2567	-0.0020	5.5
b_3	478.3388	478.3468	-0.0080	17
b_4	535.3602	535.3551	0.0051	9.5
b_5	653.2695	653.2658	0.0037	5.7

The calculated error for **4** was within 17 ppm.

[1] M. A. Seeliger, M. Young, M. N. Henderson, P. Pellicena, D. S. King, A. M. Falick, and J. Kuriyan, *Protein Sci.*, 2005, **14**, 3135-3139.

[2] R. Hili, V. Rai, and A. K. Yudin, *J. Am. Chem. Soc*., 2010, **132**, 2889-2891.