Substrate Specificity of Cytoplasmic N-Glycosyltransferase

Supplementary Information

Andreas Naegeli[1], Gaëlle Michaud[2], Mario Schubert[3], Chia-Wei Lin[1], Christian Lizak[3,4], Tamis Darbre[2], Jean-Louis Reymond[2], and Markus Aebi[1]

[1] Institute of Microbiology, Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland

[2] Department of Chemistry and Biochemistry, University of Berne, 3012 Berne, Switzerland

[3] Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland

[4] present address: Department of Biochemistry and Molecular Biology, Centre for Blood Research, University of British Columbia, Vancouver, BC V6T1Z3, Canada

---

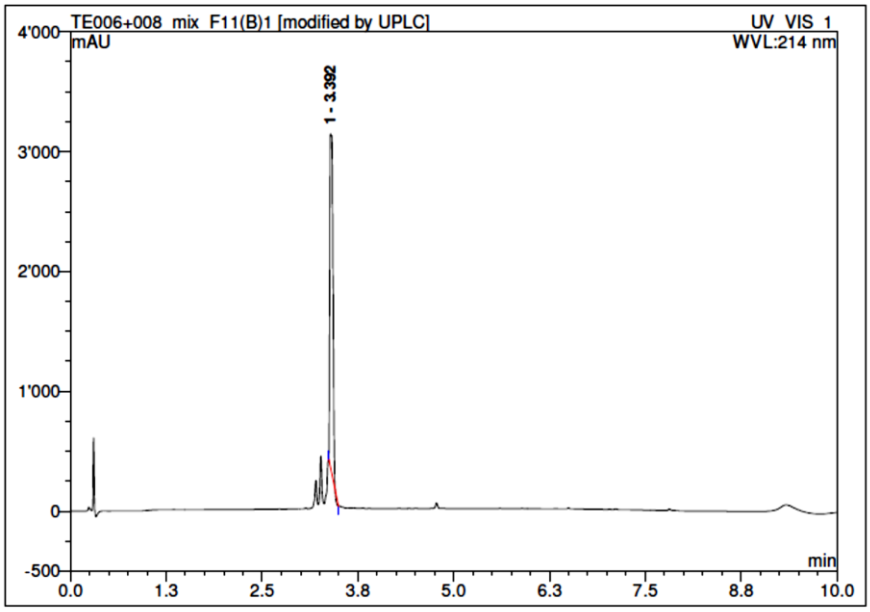
SUPPEMENTARY EXPERIMENTAL PROCEDURES

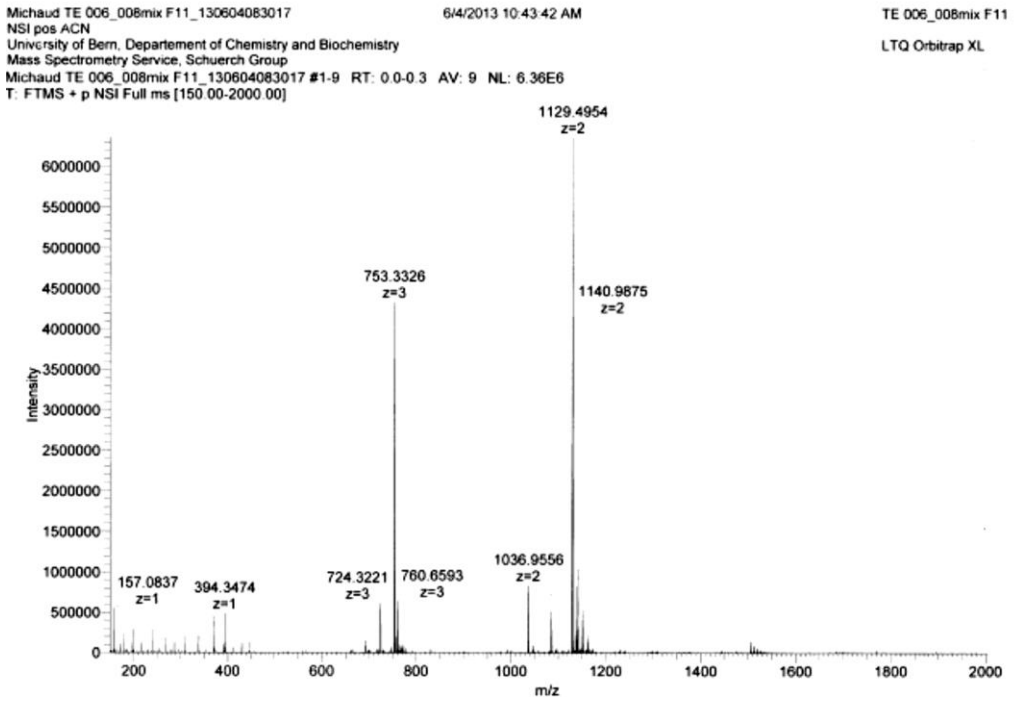
Synthesis of peptide 5-CF-SGAMGDKSVANATYSLALGS-NH_2

All reagents were either purchased from Sigma-Aldrich, Fluka, Acros Organics or Alfa Aesar. 5-Carboxyfluorescein (5-CF) was purchased either from Novabiochem (Switzerland) or from OChem Incorporation (US). The amino acid building blocks and PyBOP were purchased from Advanced ChemTech (US) or Novabiochem (Switzerland). Amino acids were used as the following derivatives: Fmoc-Ser-OH, Fmoc-Gly-OH, Fmoc-Leu-OH, Fmoc-Ala-OH, Fmoc-Tyr(OtBu)-OH, Fmoc-Thr(OtBu)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Val-OH, Fmoc-Lys(Boc)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Met-OH. Fmoc-protected TentaGel S RAM resin (loading: 0.24 $mmol \cdot g^{-1}$) was purchased from Rapp Polymere (Germany). All solvents used in solid-phase peptide synthesis were bought in p.a quality. Analytical RP-UPLC was performed in Dionex ULTIMATE 3000 Rapid Separation LC System (ULTIMATE-3000RS diode array detector) using Dionex Acclaim® RSLC 120 C18 column (2.2 µm, 120 Å, 3.0 x 50 mm, flow 1.2 $ml \cdot min^{-1}$). Compounds were detected by UV absorption at 214 nm. Data recording and processing was performed with Dionex Chromeleon Management System Version 6.80 (analytical RP-HPLC). Preparative RP-HPLC was performed with a Waters PrepLC4000 chromatography system using a Dr. Maish GmbH Reprospher Column (C18-DE, 5 µm, 100 x 30 mm, pore size 100 Å, flow rate of 60 $ml \cdot min^{-1}$). Compounds were detected by UV absorption at 214 nm using a Waters 486 Tunable Absorbance detector. All RP-HPLC were using HPLC-grade acetonitrile and Milli-Q deionized water. The elution solutions were: **A** H_2O with 0.1% TFA; **B** H_2O/ACN

(50:50); **C** H_2O/ACN (10:90) with 0.1% TFA; **D** H_2O/ACN (40:60) with 0.1% TFA. MS spectra, recorded on a Thermo Scientific LTQ OrbitrapXL were provided by the Mass Spectrometry service of the Department of Chemistry and Biochemistry at the University of Berne.

From Tenta Gel S RAM® resin (500 mg, loading: 0.24 mmol·g^{-1}), the purest fraction was isolated as a yellow foamy solid after preparative RP-HPLC purification (9.3 mg, 3.3 % yield). Gradient used for preparative RP-HPLC is A/D = 100/0 to 60/40 in 45 min, A/D 60/40 to 0/100 in 10 min with a flow rate of 40 ml·min^{-1}. Analytical. UHPLC: t_R = 3.39 min (A/C = 100/0 to 0/100 in 7.5 min, flow rate 1.2 ml·min^{-1}). HRMS (ESI$^+$) calc. for $C_{100}H_{141}N_{23}O_{35}S$ $[M+H]^+$: 2256.98; obsd: 1129.49 (z=2), 753.33 (z=3).





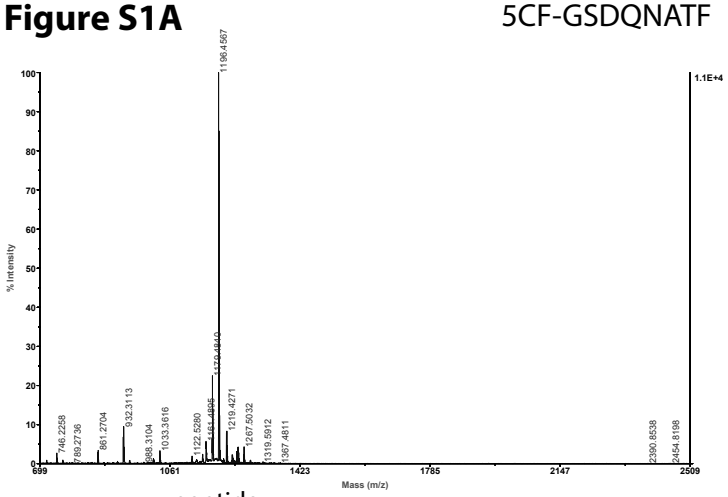
SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Mass spectrometric analysis of peptide substrate specificity of ApNGT (A) MALDI mass spectra of synthetic substrate peptides before and after incubation with ApNGT and UDP-Glc. Peaks corresponding to glycopeptides were detected for peptides 5CF-GSDQNATF ($[M+H]^+$=1358.49) and 5CF-GSDQhSATF ($[M+H]^+$=1345.49). Glycosylation of peptide 5CF-GSDQQATF could not be demonstrated by MALDI-MS but LC-ESI-MS/MS analysis (panel (C)) showed its glycosylation. No glycopeptide product could be detected for the other peptides. (B) Side chain structures of unnatural amino acids used (C) MS2 spectra from LC-ESI-CID-MS/MS analysis of the glycosylated peptides 5CF-GSDQQATF and 5CF-GSDQhSATF demonstrating glycosylation of glutamine and homoserine by ApNGT. The hashtag (#) marks ions resulting from precursor glycopeptide neutral loss.

Figure S2: GT41 Alignment Complete alignment of 3 NGT sequences (from *A. pleuropneumoniae*, *H. influenza*e and *Y. enterocolitica*) and 4 OGT sequences (human, *C. elegans*, *D. melanogaster*, *X. campestris*). The alignment was performed using Expresso (see experimental procedures for details). Parts of the alignment are shown in Figure 4

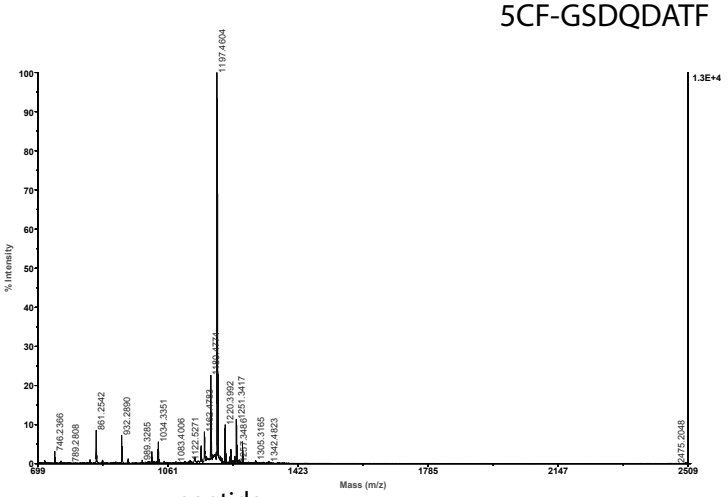
Figure S1A

5CF-GSDQNATF



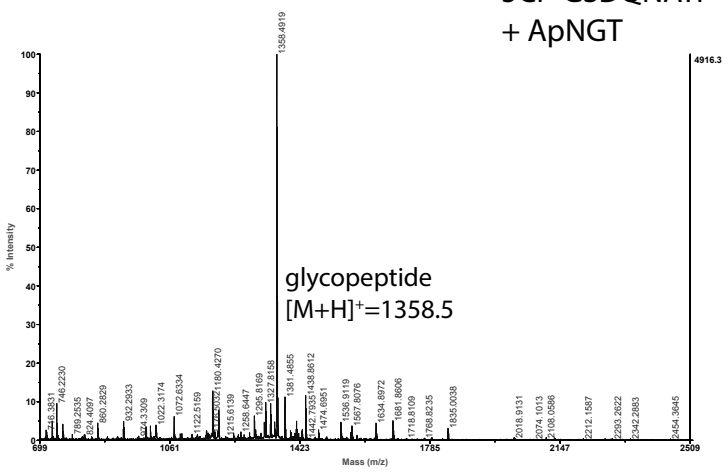
peptide
$[M+H]^+$=1196.5

5CF-GSDQDATF

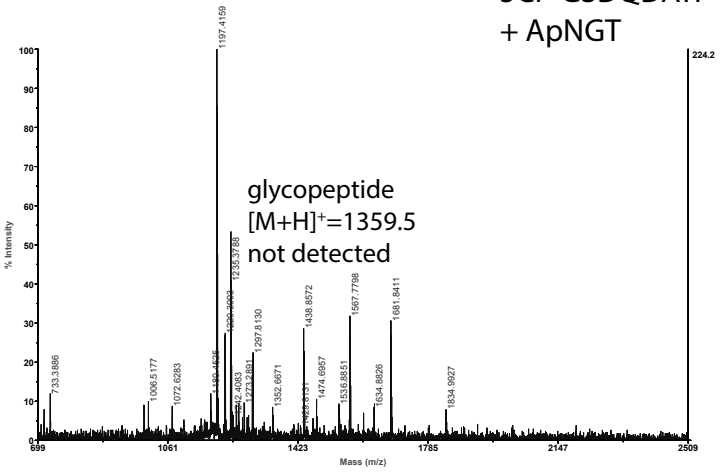


peptide
$[M+H]^+$=1197.5

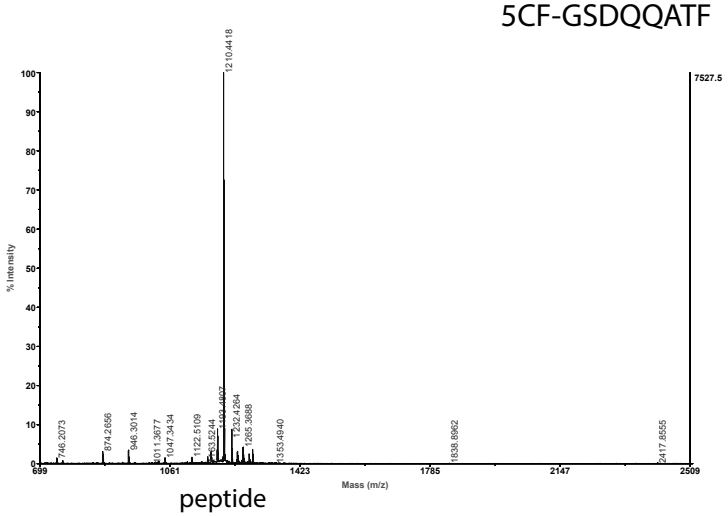
5CF-GSDQNATF
+ ApNGT



5CF-GSDQDATF
+ ApNGT

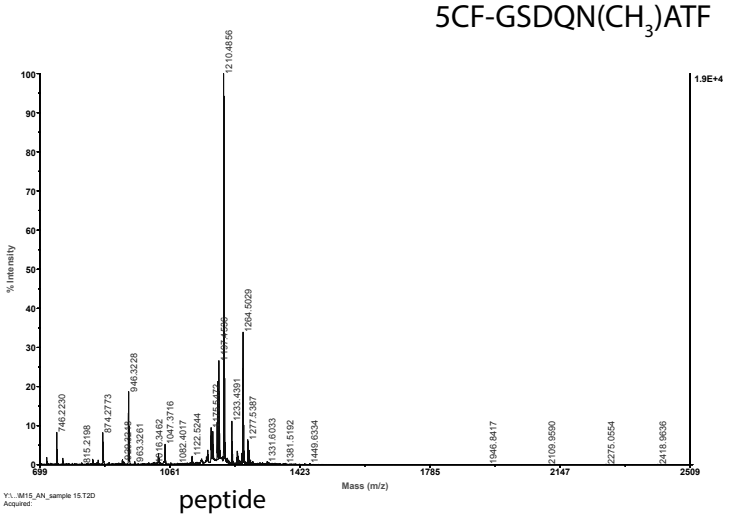


5CF-GSDQQATF



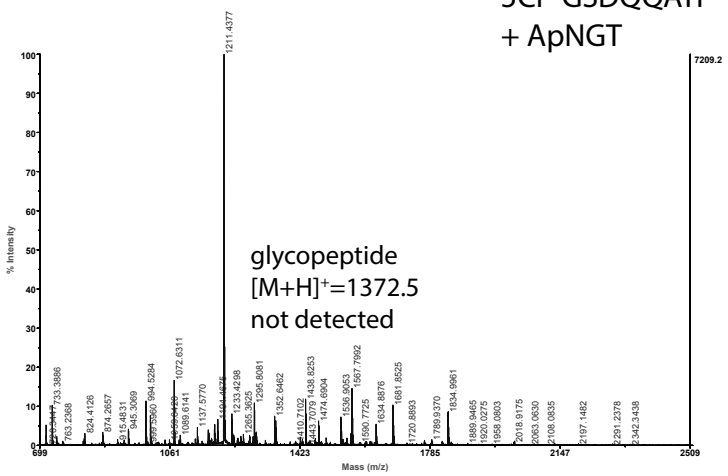
peptide
$[M+H]^+$=1210.5

5CF-GSDQN(CH_3)ATF

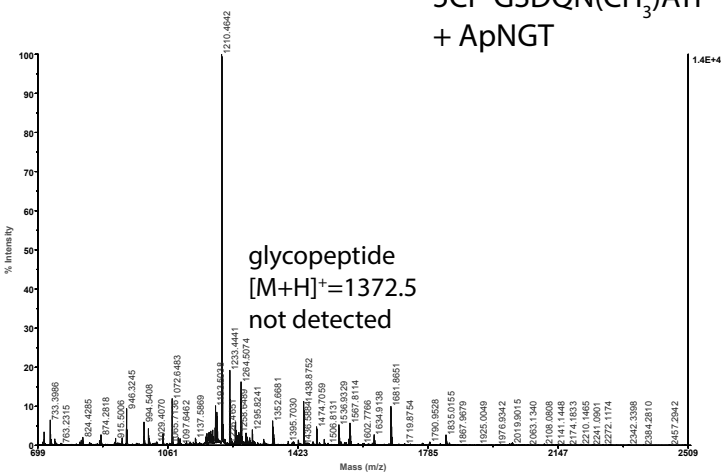


peptide
$[M+H]^+$=1210.5

5CF-GSDQQATF
+ ApNGT



5CF-GSDQN(CH_3)ATF
+ ApNGT



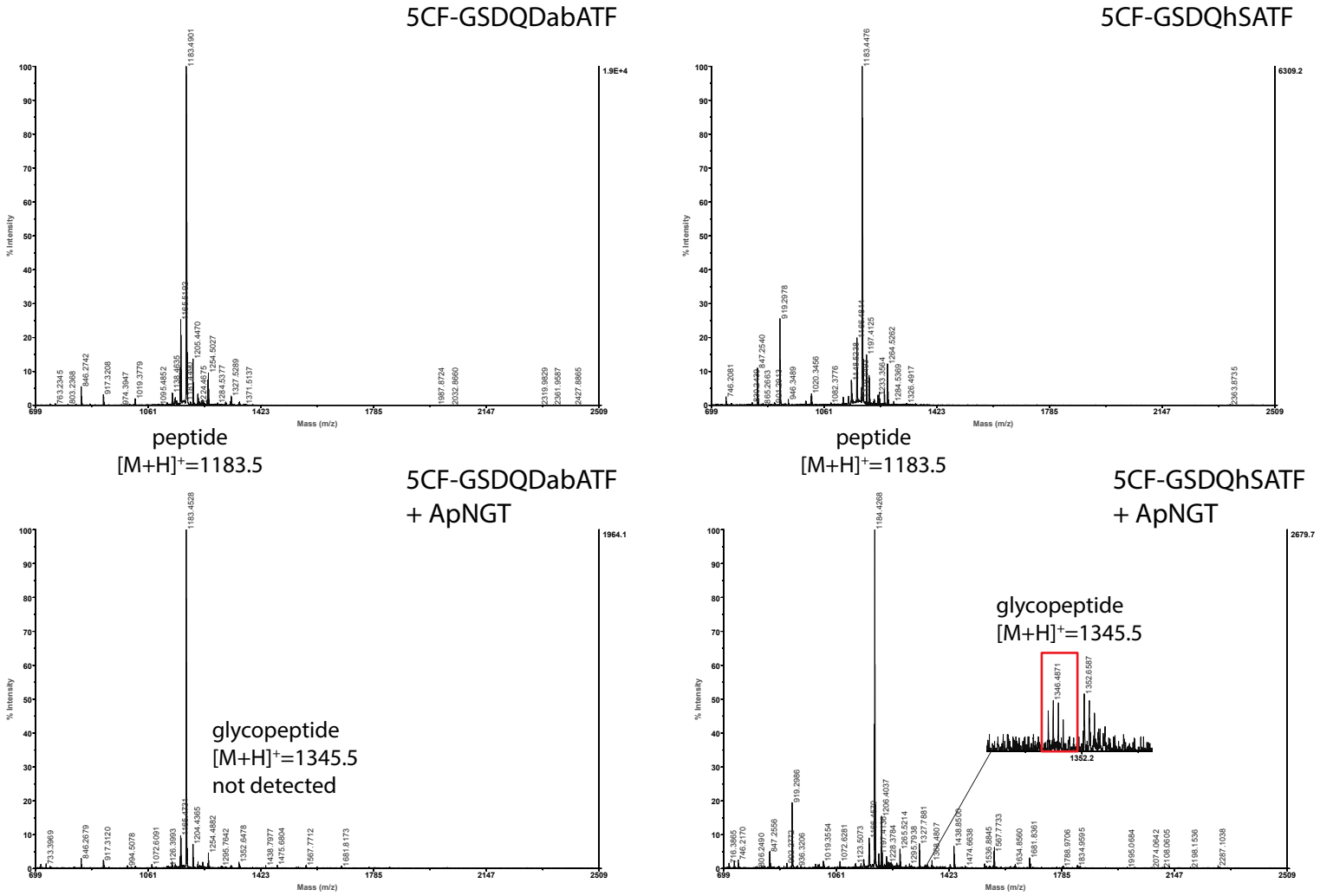
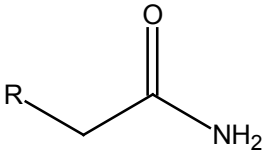
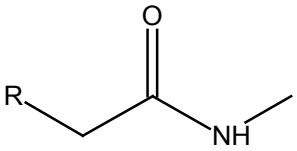


Figure S1B

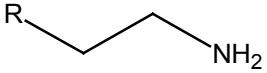
Asparagine
N



N-Methylasparagine
N(CH_3)



Diamino butyric acid
Dab



Homoserine
hS

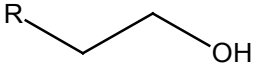
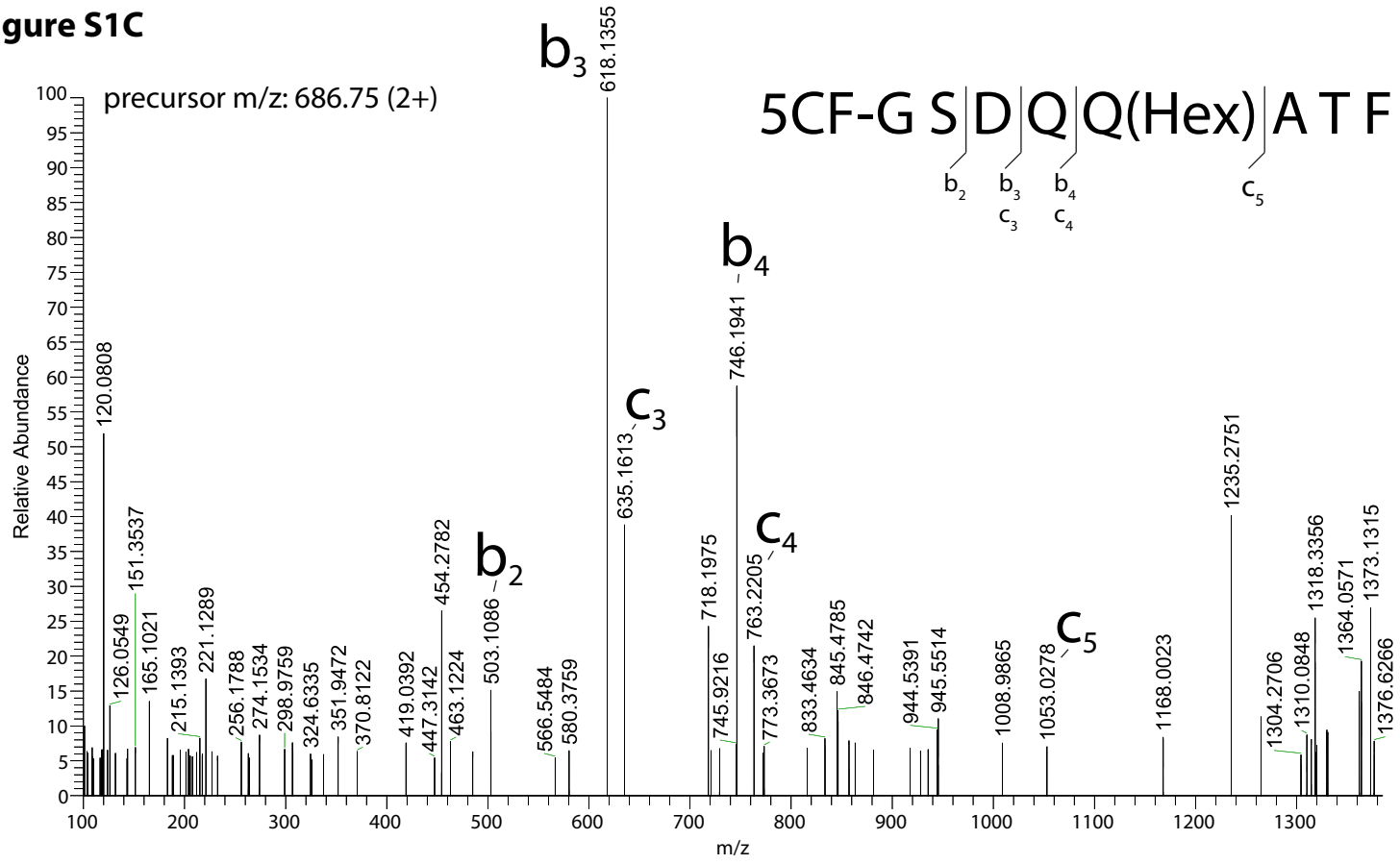


Figure S1C


precursor m/z: 686.75 (2+)
5CF-G S D Q Q(Hex) A T F
b2 b3 b3 c3 b4 c4 c5
b3 618.1355
b4 746.1941
c3 635.1613
c4 763.2205
b2 503.1086
c5 1053.0278
120.0808
126.0549
151.3537
165.1021
215.1393
221.1289
256.1788
274.1534
298.9759
324.6335
351.9472
370.8122
419.0392
447.3142
454.2782
463.1224
566.5484
580.3759
718.1975
745.9216
773.3673
833.4634
845.4785
846.4742
944.5391
945.5514
1008.9865
1168.0023
1235.2751
1304.2706
1310.0848
1318.3356
1364.0571
1373.1315
1376.6266
Relative Abundance
m/z

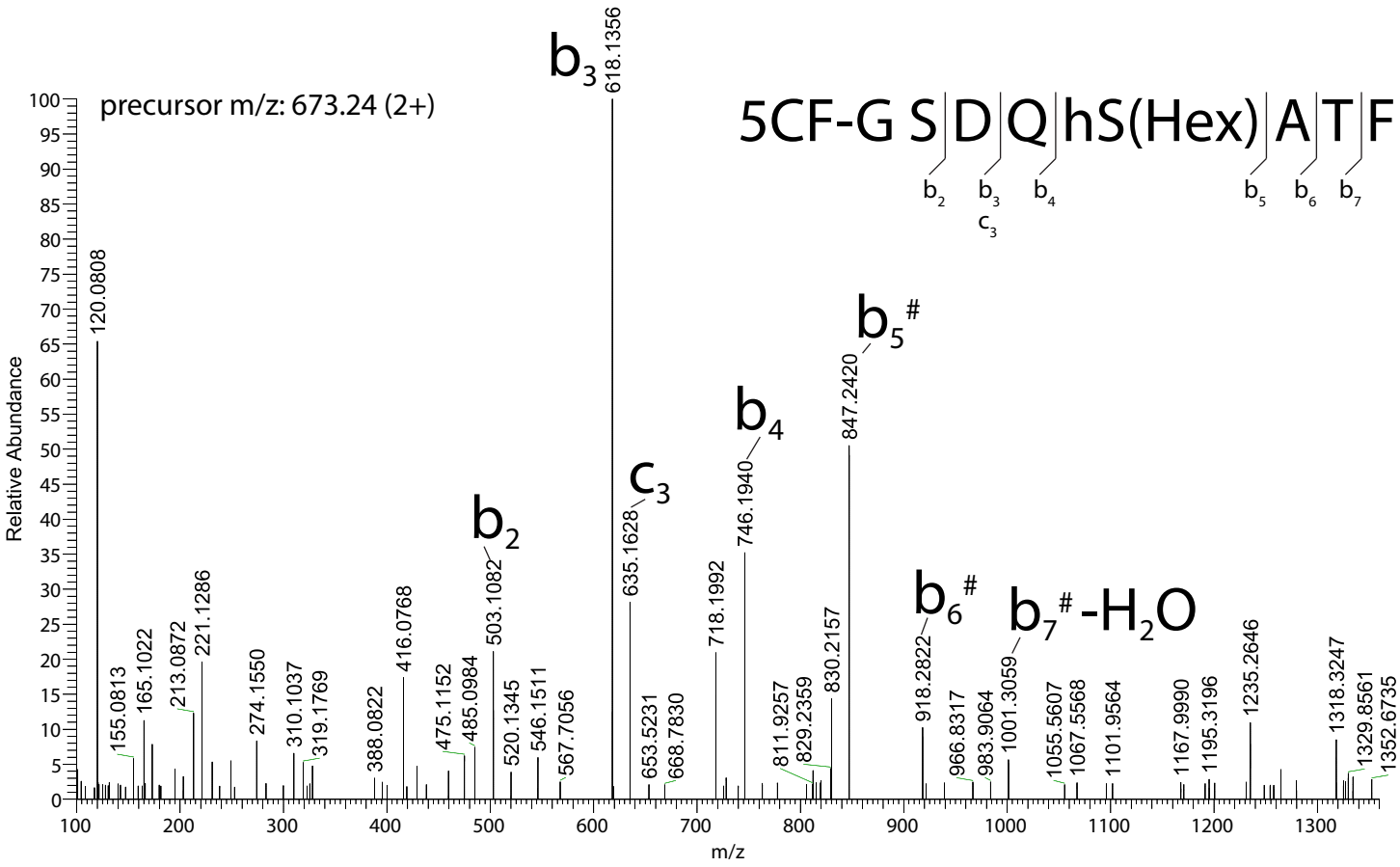
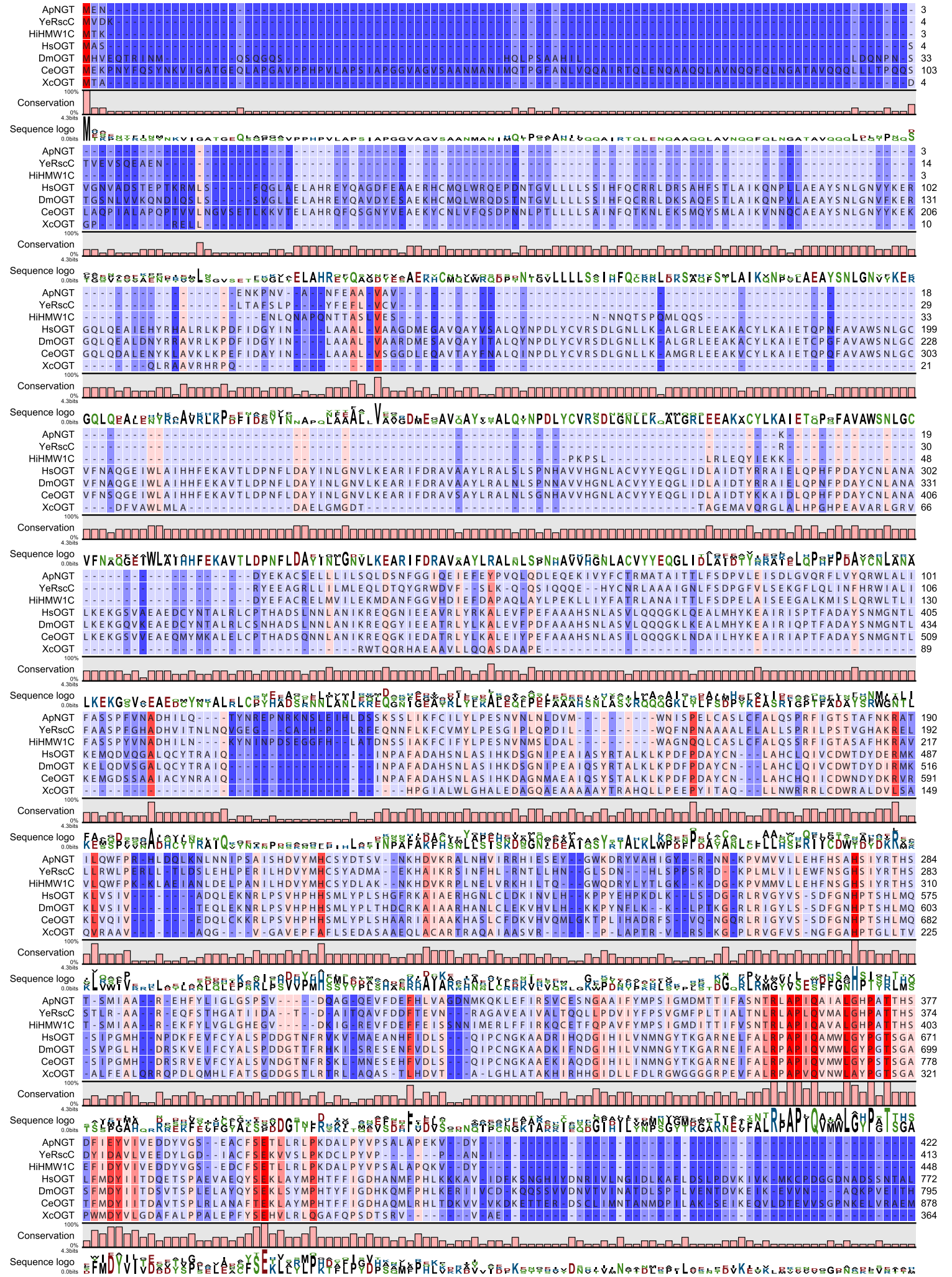
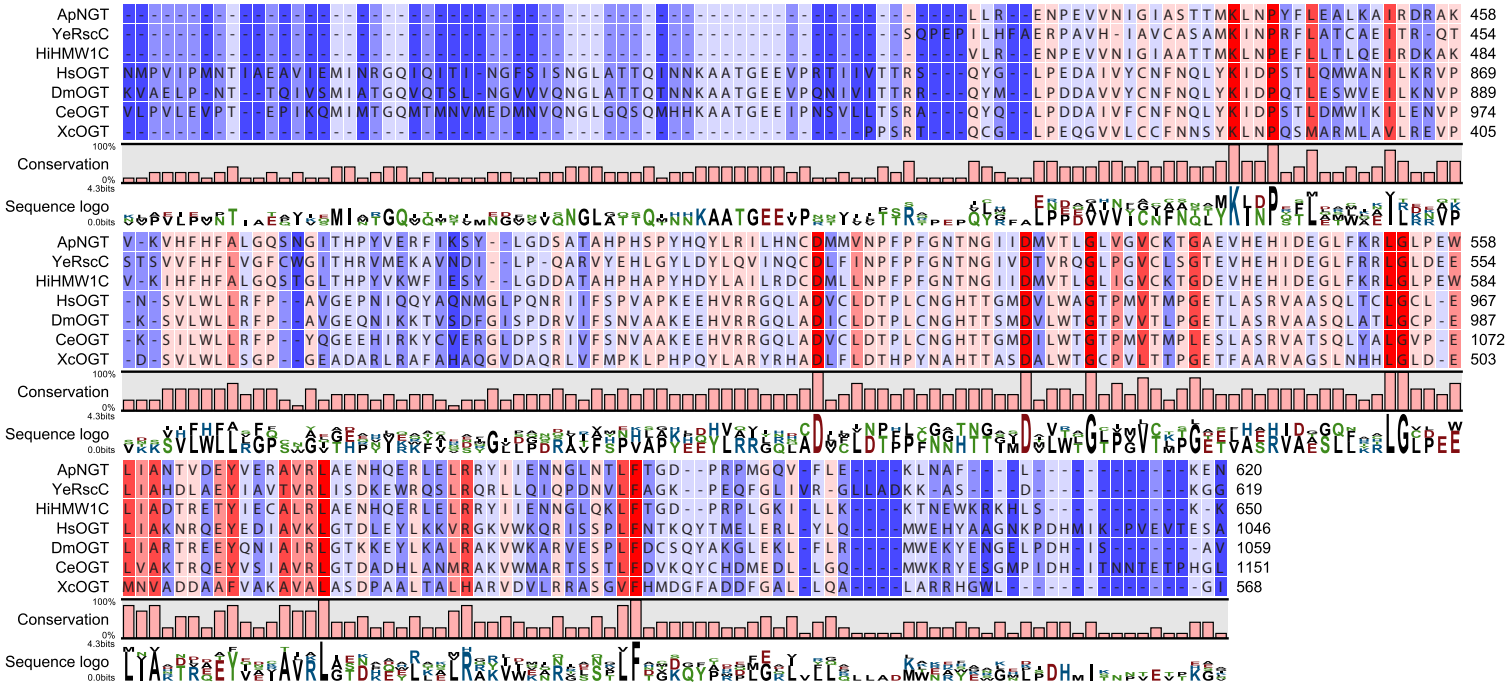

precursor m/z: 673.24 (2+)
5CF-G S D Q hS(Hex) A T F
b2 b3 b3 c3 b4 b5 b6 b7
b3 618.1356
b5# 847.2420
b4 746.1940
c3 635.1628
b2 503.1082
b6# 918.2822
b7#-H2O 1001.3059
120.0808
155.0813
165.1022
213.0872
221.1286
274.1550
310.1037
319.1769
388.0822
416.0768
475.1152
485.0984
520.1345
546.1511
567.7056
653.5231
668.7830
718.1992
811.9257
829.2359
830.2157
966.8317
983.9064
1055.5607
1067.5568
1101.9564
1167.9990
1195.3196
1235.2646
1318.3247
1329.8561
1352.6735
Relative Abundance
m/z

Figure S2


ApNGT
YeRscC
HiHMW1C
HsOGT
DmOGT
CeOGT
XcOGT
Conservation
Sequence logo


ApNGT
YeRscC
HiHMW1C
HsOGT
DmOGT
CeOGT
XcOGT
Conservation
Sequence logo
458
454
484
869
889
974
405
558
554
584
967
987
1072
503
620
619
650
1046
1059
1151
568