

SCIENTIFIC REPORTS



OPEN

Improvement of the glycoproteomic toolbox with the discovery of a unique C-terminal cleavage specificity of flavastacin for N-glycosylated asparagine

Alexander Pralow¹, Marcus Hoffmann¹, Terry Nguyen-Khuong¹, Erdmann Rapp^{1,2} & Udo Reichl^{1,3}

To determine all potential N-glycosylation sites of a glycoprotein, one central aspect of every bottom-up N-glycoproteomic strategy is to generate suitable N-glycopeptides that can be detected and analyzed by mass spectrometry. Specific proteases, such as trypsin, bear the potential of generating N-glycopeptides that either carry more than one N-glycosylation site or are too long to be readily analyzed by mass spectrometry—both due to the lack of tryptic cleavage sites near the N-glycosylation site. Here, we present a newly identified cleavage specificity of flavastacin, a protease from *Flavobacterium meningosepticum*, which - up to now - was only reported to cleave peptide bonds N-terminal to aspartic acid residues. In contrast to literature, we could not confirm this N-terminal specificity of flavastacin for aspartic acid. However, for the first time, we show a unique cleavage specificity of flavastacin towards the C-terminus of N-glycosylated asparagine residues. Implemented in an N-glycoproteomic workflow the use of flavastacin can thus not only render data analysis much easier, it can also significantly increase the confidence of MS-based N-glycoproteomic analyses. We demonstrate this newly discovered specificity of flavastacin by in-depth LC-MS(/MS) analysis of complex-type glycosylated human lactotransferrin and bovine serum albumin peptides and N-glycopeptides that were generated by trypsin and flavastacin digestion. Following to this work, further elucidation of the efficiency, specificity and mode of action of flavastacin is needed, but we believe that our discovery has great potential to facilitate and improve the characterization of N-glycoproteomes.

Protein glycosylation is a co-/post-translational modification involved in several key biological functions¹. It is becoming increasingly evident that aberrant glycosylation is associated with cancer², inflammatory diseases and infectious diseases³. In addition, there is a rapidly expanding group of rare genetic, metabolic disorders that are due to defects in glycosylation, the so-called congenital disorders of glycosylation⁴. There are two main forms of protein glycosylation: N-linked and O-linked glycosylation. N-glycans are linked to the amino group of asparagine according to a specific consensus sequence (NXS/T; X ≠ P). Moreover, N-glycans are characterized by a common core-structure GlcNAc₂Man₃ (N-acetylglucosamine (GlcNAc), mannose (Man)), which can be extended to form complex-, high-mannose- or hybrid-type N-glycan structures⁵. In the past decades, the field of site-specific glycoprotein analysis methods has grown tremendously and emphasized the importance of protein glycosylation analyses with respect to micro- (different glycan structures at one glycosylation site) and macro-heterogeneity (site occupancy). So far, the method of choice for site-specific glycoproteomics is the analysis of specifically or unspecifically digested glycoproteins via liquid chromatography combined with mass spectrometry

¹Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, 39106, Magdeburg, Germany.

²glyXera GmbH, Leipziger Straße 44, 39120, Magdeburg, Germany. ³Otto-von-Guericke University, Chair of Bioprocess Engineering, Universitätsplatz 2, 39106, Magdeburg, Germany. Correspondence and requests for materials should be addressed to E.R. (email: rapp@mpi-magdeburg.mpg.de)

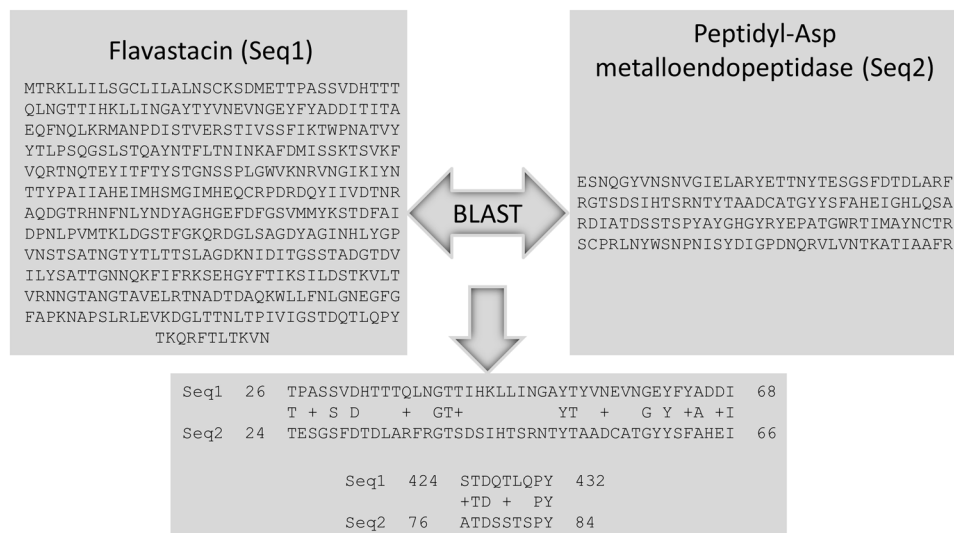


Figure 1. Comparison of flavastacin and peptidyl-Asp metalloendopeptidase. The protein sequences of flavastacin (Q47899) and peptidyl-Asp metalloendopeptidase (Q9R4J4) are from Uniprot. Both sequences were compared using the protein BLAST tool from the national center for biotechnology information (free online tool: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=blasttab). Sequences with similarities are illustrated in the lower box. The first sequence has a similarity of 11 out of 43 amino acids (23%). The second sequence has a similarity of 4 out of 9 amino acids (44%). Possible conservative amino acid changes are indicated with “+”.

(LC-MS). In addition, a glycopeptide enrichment can be performed prior to analysis, i.e. by hydrophilic interaction chromatography⁶. With regard to MS, different fragmentation strategies tackle various challenges in the field of proteomics and glycoproteomics. Lower energy collision-induced dissociation (CID) is able to generate B- and Y-glycomoiety fragment ion series of glycopeptides⁷. It allows the annotation of the glycan composition and the calculation of the peptide mass, but often lacks b- and y-peptidomoiety fragment ion series⁸ of the glycopeptides for sequence verification. Higher energy collisional dissociation (HCD) allows the adjustment of the normalized collision energy and generates peptide-specific b- and y-ions. However, due to the neutral loss of the glycan moiety, B- and Y-ion series are underrepresented^{9,10}.

In addition to tryptic digestion of glycoproteins, further sequential treatment with other proteases is used to overcome low charge density and sequence constraints of glycopeptides with a large peptide moiety. The endoprotease AspN is a zinc metalloendoprotease produced in *Pseudomonas fragi* (Boehringer Ingelheim, Uniprot: Q9R4J4), which selectively cleaves peptide bonds N-terminal to aspartic acid¹¹. AspN is also known for N-terminal cleavage at cysteine and glutamic acid^{12,13}. Its primary use in glycoproteomic experiments so far, has involved the cleavage of deamidated asparagine after N-glycan release by peptide N-glycosidase F (PNGaseF) to assess N-glycan presence and location¹⁴. Flavastacin (New England Biolabs, Uniprot: Q47899), which is produced in *Flavobacterium meningosepticum*, has been described to behave similar to the AspN from *Pseudomonas fragi*^{15,16}. Therefore, is also called AspN, despite its quite different amino acid sequence and thus, protein identity. However, both proteins belong to the family of metalloendoproteases. A BLAST search comparing these two proteins shows no overlapping sequences and only two short segments with quite low similarity (see Fig. 1). Thus, the two sequences cannot be aligned, which renders common protein functions unlikely.

This work describes a new and unique protease specificity of flavastacin for the C-terminus of N-glycosylated asparagine. In contrast to literature^{15,16}, no specificity for the N-terminus of aspartic acid was observed. Subsequent analysis of this unexpected phenomenon via *de-novo* sequencing of the resulting N-glycopeptide fragment ion spectra led to the discovery of this previously unknown cleavage specificity of flavastacin. Our findings provide investigators with a new tool for targeted N-glycoprotein digestion to overcome common problems in N-glycoproteomics, like large N-glycopeptides with too many amino acids for proper LC-MS measurements, or multiple N-glycosylation sites within one N-glycopeptide.

Materials and Methods

Chemicals. The proteins bovine serum albumin (BSA; A3912–100G) and lactotransferrin from human milk (hLTF; L4894-5MG) were purchased from Sigma-Aldrich. Enzymes used for digestion were trypsin (Trypsin Sequencing Grade Modified; V5111) from Promega and endoprotease AspN (AspN; P8104S) from New England Biolabs (Flavastacin, purified host cell protein, Uniprot Q47899; see the corresponding SDS-gel for purity of AspN (Flavastacin) in Supplementary Figure 1). All solvents for LC were MS grade. All buffer and solutions were prepared with deionized and purified water (dH₂O) using a Milli-Q water purification system (18.2 MΩ · cm⁻¹ at 25°C, total organic carbon of 3 ppb) from Merck Millipore. For LC-MS solvents, water was further purified using the LC-Pak Polisher from Merck Millipore.

N-Glycosylation Site	Experimental/Observed mass [Da]	Theoretical mass [Da]	Δ Mass [Da]	Peptide	Position	Modifications	RT [min]
156	490.2640	490.2660	0.002	(R)PFLN(W)	153–156		24.73
	860.4960	860.4990	0.003	(G)TLRPFLN(W)	150–156		25.07
	1127.6620	1127.6570	−0.005	(V)PIGTLRPFLN(W)	147–156		40.09
	1226.7230	1226.7260	0.003	(N)VPIGTLRPFLN(W)	146–156		41.54
	1755.9520	1755.9540	0.002	(R)TAGWNVPIGTLRPFLN(W)	141–156		42.17
	1340.7680	1340.7680	0.000	(W)NVPIGTLRPFLN(W)	145–156		42.93
497	856.4950	856.4961	−0.001	(N)IPM(−48)GLLFN(Q)	490–497	M(CAM) (−105)	30.11
	506.2990	506.2970	−0.002	(G)LLFN(Q)	494–497		32.10
	563.3210	563.3190	−0.002	(M)GLLFN(Q)	493–497		36.07
	1385.7223	1385.7246	−0.002	(R)TAGWNIPM(−48)GLLFN(Q)	485–497	M(CAM) (−105)	39.58
	920.4930	920.4910	−0.002	(N)IPMGLLFN(Q)	490–497	MSO	39.95
	904.4950	904.4960	0.001	(N)IPMGLLFN(Q)	490–497		45.56
	791.4110	791.4120	0.001	(I)PMGLLFN(Q)	491–497		47.45
	694.3590	694.3590	0.000	(P)MGLLFN(Q)	492–497		53.33

Table 1. Manually annotated N-glycopeptide sequences of human lactotransferrin (hLTF) from MS(/MS) spectra of nanoRP-LC-ESI-OT-MS² (HCD) measurements after sequential digestion of hLTF with trypsin and flavastacin. CAM - carbamidomethylation; MSO - methionine S-oxidation.

Proteolytic Digestion via Filter-Aided Sample Preparation (FASP). BSA and hLTF were digested proteolytically using a modified version of the filter-aided sample preparation (FASP) method¹⁷. Briefly, 100 μ g of each protein were applied to a filter unit (Nanosep[®] Omega[™] with polyethersulfone membrane, molecular weight cut-off 10 kDa; PALL Life Sciences). Samples were treated with urea buffer_(Tris-HCl) (8 M urea in 0.1 M Tris-HCl_(aq) pH 8.5; AppliChem), followed by reduction with DL-dithiothreitol (40 mM DTT, Sigma-Aldrich), and alkylation with iodoacetamide (55 mM IAA, Sigma-Aldrich) – each dissolved in 50 mM ammonium bicarbonate_(aq) (ABC buffer_(aq), Sigma-Aldrich). Each filter unit was washed three times with urea buffer_(Tris-HCl) and three times with ABC buffer_(aq). Proteins were digested proteolytically with trypsin using an enzyme/protein ratio of 1:30 (w/w). Samples were incubated overnight at 37°C and 350 rpm using a temperature controlled incubator (Titramax 1000 + Inkubator 1000, Heidolph). Digests were collected by centrifugation. Filter units were washed twice, first using 50 μ L ABC buffer_(aq) with 5%_(v/v) acetonitrile (ACN), then using 50 μ L dH₂O; in between samples were centrifuged. The flow through was kept along with the digest, in order to be dried by vacuum centrifugation.

After tryptic digestion, approximately 20 μ g peptides were reconstituted in 20 μ L 1x AspN reaction buffer (New England Biolabs; 50 mM Tris-HCl, 2.5 mM Zinc Sulfate, pH 8.0). Afterwards AspN (Flavastacin) was added (enzyme/protein ratio 1:20) to the peptide solution and incubated overnight at 37°C as recommended by the supplier. The enzyme reaction was stopped via centrifugation through a filter unit (same as described above). The flow through (tryptic digests of BSA and hLTF, as well as the sequential digests of BSA and hLTF with trypsin and AspN (Flavastacin) was dried by vacuum centrifugation and reconstituted in 0.1%_(v/v) trifluoroacetic acid_(aq) (TFA; Thermo Fisher Scientific) prior to LC-MS(/MS) measurements.

LC-MS(/MS) Measurement. The LC system was an UltiMate 3000 Rapid Separation LC system from Thermo Fisher Scientific. Samples (\approx 500 ng) were loaded isocratically on a trap column (Acclaim PepMap[®] 100, 100 μ m \times 2 cm nanoViper C18, 5 μ m, 100 Å, Thermo Fisher Scientific) via 100% loading buffer A (98%_(v/v) dH₂O, 2%_(v/v) ACN, 0.05%_(v/v) TFA) with a flow rate of 7 μ L/min within the first 5 min. Afterwards, the loaded trap column was switched in line with the separation column (Acclaim PepMap[®]RSLC, 75 μ m \times 25 cm nanoViper C18, 2 μ m, 100 Å, Thermo Fisher Scientific), with a nano flow rate of 0.3 μ L/min of 4% nano buffer B (10%_(v/v) dH₂O, 10%_(v/v) trifluoroethanol (TFE), 80%_(v/v) ACN, 0.1%_(v/v) formic acid (FA)) and nano buffer A (98%_(v/v) dH₂O, 2%_(v/v) ACN, 0.1%_(v/v) FA). The separation was performed by a multi-step binary nano A/B gradient: 4–55% nano buffer B till 80 min, 55–90% nano buffer B till 100 min, 90% nano buffer B till 110 min and 4% nano buffer B 110–150 min.

The eluting peptides were measured on an LTQ Orbitrap Elite mass spectrometer from Thermo Fisher Scientific using a Nanospray Flex[™] source in positive ionization mode with a capillary voltage of −2.7 kV. Peptides and glycopeptides were fragmented using HCD with normalized collision energy of 35 with an activation time of 0.1 ms. The five most intense precursor ions with a charge state >1 were chosen for fragmentation. The recorded mass range for MS was 350–2000 m/z and for MS/MS 150–2000 m/z.

Data Analysis. The complete MS(/MS) data for hLTF (Trypsin + AspN (Flavastacin)) were analyzed manually using Xcalibur (Version 2.2, Qual Browser, Thermo Fisher Scientific). The first step was the recognition of glycopeptide related fragment ion spectra due specific B-ions “oxonium ions”. Afterwards, the peptide mass was presumed by a specific fragmentation pattern: [peptide − NH₃ + H⁺]; [peptide + H⁺] and [peptide + GlcNAc + H⁺]. Using ExPASy Findpept (free online tool: web.expasy.org/findpept) the presumed peptide mass was screened against an unspecific digestion of hLTF (Uniprot; P02788) with a maximum mass tolerance of 0.02 Da, carbamidomethylation of cysteine and the oxidation of methionine. The presumed peptide sequence was

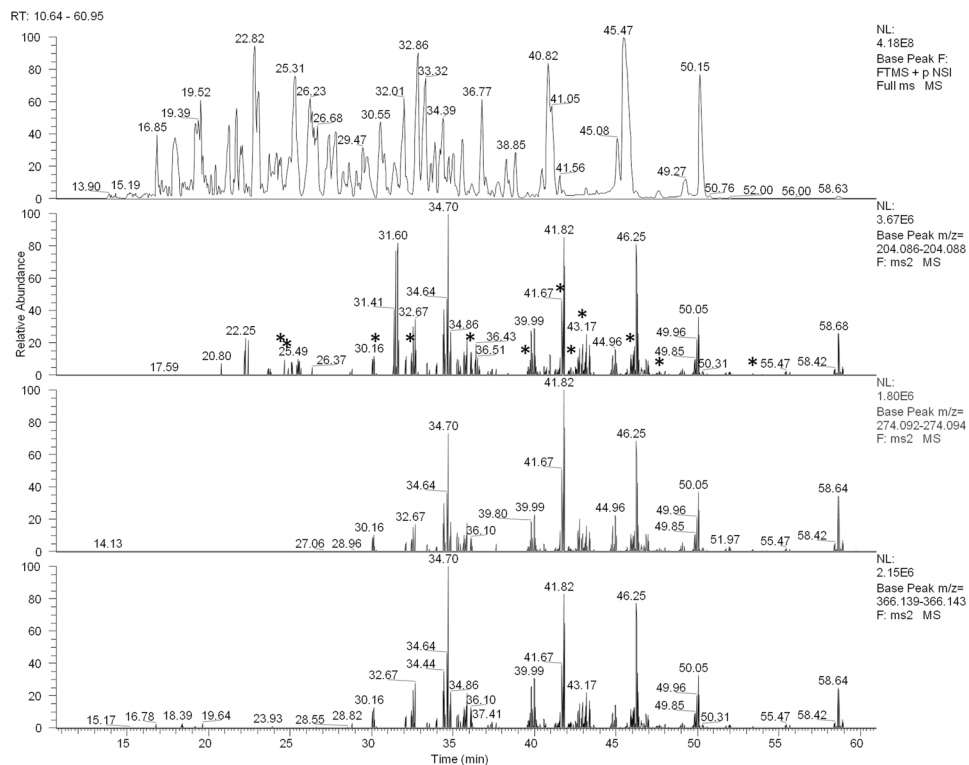


Figure 2. Base peak ion-chromatogram (BPC) and oxonium ion related extracted ion-chromatograms (EIC) of MS(/MS) spectra of nanoRP-LC-ESI-OT-MS² (HCD) measured hLTF after sequential digestion with trypsin and flavastacin. Starting from the top: BPC of MS spectra (grey), EIC of MS/MS spectra of HexNAc within the m/z range 204.086–204.088 [M + H]⁺ (red), EIC of MS/MS spectra of NeuAc-H₂O within the m/z range 274.092–274.094 (green), EIC of MS/MS spectra of HexHexNAc within the m/z range 366.139–366.143 (blue). The ion-chromatograms are illustrated in the time range 10.64–60.95 min. The accepted mass error of the EIC of the specific oxonium ions is 5 ppm. Signals listed in Table 1 are marked using asterisks.

furthermore *in-silico* fragmented using *MS-Product* (free online tool: prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct). The b- and y-ions were compared with the corresponding ions in the MS/MS fragment ion spectrum with a maximum mass tolerance of 0.02 Da to validate the presumed peptide sequence. The mass difference between the peptide mass and the precursor mass was used to predict the N-glycan composition using *ExPASy GlycoMod* (free online tool: web.expasy.org/glycomod).

The MS(/MS) data from BSA and hLTF were imported into *Proteome Discoverer* (Version 1.4, Thermo Fisher Scientific) and searched against UniProt-KB/SwissProt database (542258 sequences; downloaded January, 2014) using *MASCOT* (Version 2.5, Matrix Science). The MS(/MS) data were screened against an unspecific *in-silico* digestion of the mammalian taxonomy database with the fixed modification of cysteine with carbamidomethyl, variable deamidation of asparagine, and variable oxidation of methionine. The precursor ion mass tolerance was set to 5 ppm and the fragment ion mass tolerance to 0.02 Da. The protein relevance threshold was set to 20 and the peptide cut off score to 10. The target false discovery rate for peptide hits was set to 0.01 (strict setting, and to 0.05 as relaxed setting).

Data availability statement. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Results and Discussion

Every N-glycoproteomic analysis workflow consists of numerous parameters to be optimally adjusted. In particular, the design of proteolytic digestion using sequential digestion steps with a selection of specific enzymes is an important step to overcome common problems such as too large N-glycopeptides (with low charge density and/or sequence constraints) and N-glycopeptides with multiple glycosylation sites. Here, we present a new approach for the proteolytic digest of glycoproteins by using flavastacin, a protease that we found to cleave specifically at the C-terminus of N-glycosylated asparagine. The glycoprotein hLTF and the non-glycosylated protein BSA were used as model proteins to demonstrate this newly identified cleavage specificity. According to the manufacturer's recommendation, flavastacin works only on peptides smaller than 50 amino acids. Therefore, hLTF and BSA were first treated with trypsin before flavastacin was added (see Materials and Methods).

The hLTF is a well-characterized glycoprotein present in human milk, saliva, tears, nasal secretions and other body fluids¹⁸ that contains three potential N-glycosylation sites (N₁₅₆, N₄₉₇, and N₆₄₂). The sites N₁₅₆ and

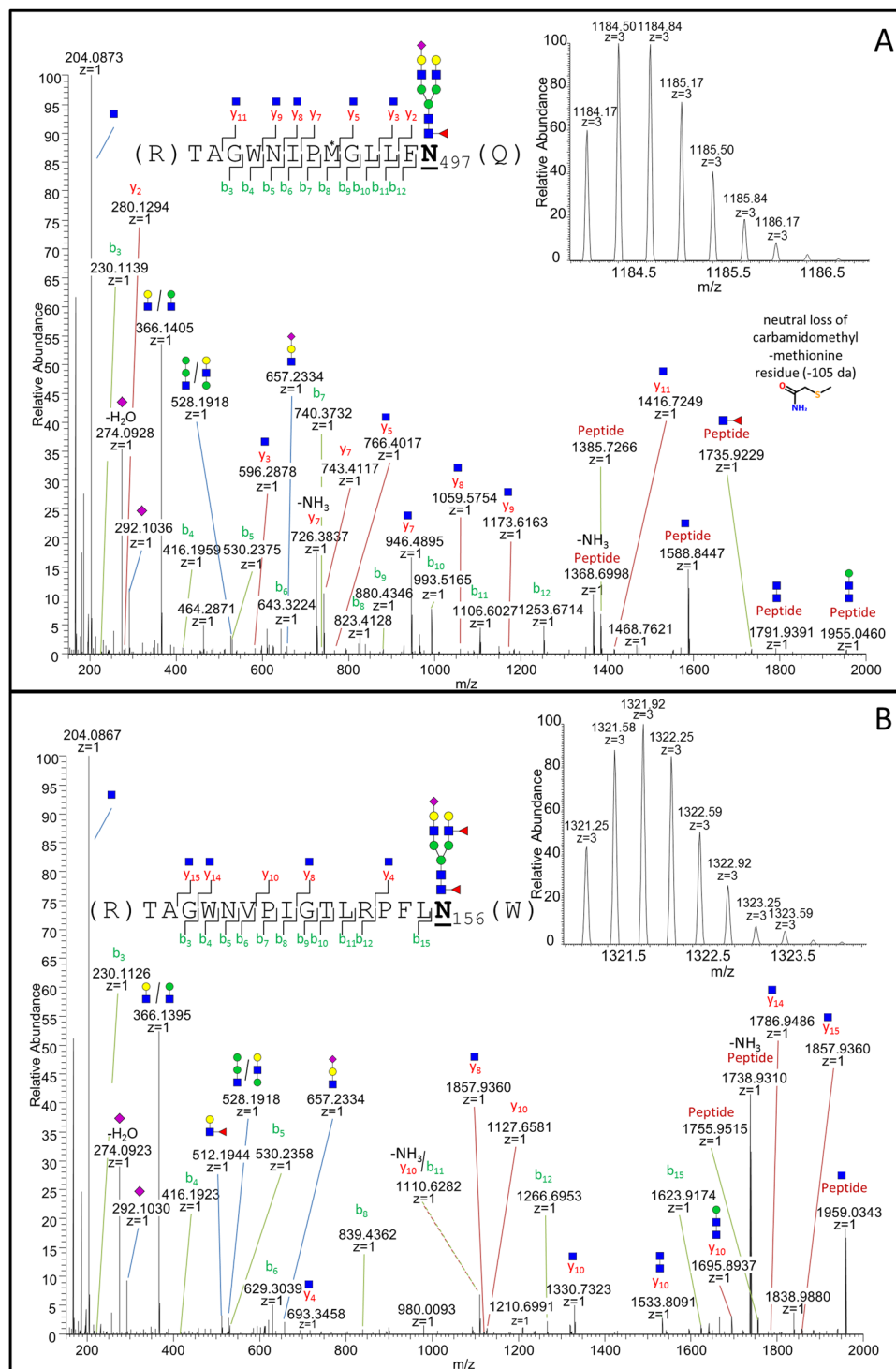


Figure 3. Fragment ion spectra of nanoRP-LC-ESI-OT-MS² (HCD) measured hLTF N-glycopeptides after sequential digestion with trypsin and flavastacin. **(A)** For site N₄₉₇ the fragment ion spectrum of the N-glycopeptide sequence TAGWNIPM*GLLFN₄₉₇ with the N-linked glycan Hex₂HexNAc₄Hex₁NeuAc₁ and the corresponding precursor ion m/z 1184.1659 [M + 3H]³⁺ is shown. The “*” indicates the carbamidomethylation of methionine. **(B)** For site N₁₅₆ the fragment ion spectrum of the N-glycopeptide sequence TAGWNVPIGTLRPFLLN₁₅₆ with the N-linked glycan Hex₂HexNAc₄Hex₁NeuAc₁ and the corresponding precursor ion m/z 1321.2479 [M + 3H]³⁺ is shown. N-glycan structures and oxonium ions are illustrated according to CFG nomenclature²². The b-ion series are highlighted in green, the y-ion series in red and the B-ions in blue. The isotopic patterns of the precursors are shown at the upper right corner.

Sequence	# PSM	Modifications	MH + [Da]	IonScore	ΔM [Da]
(K)TVMENFVAFVVDK(C)	2	N5(Deamidated)	1400.67510	107	-0.002
(R)MPC TEDYLSLILNR(L)	2	C3(Carbamidomethyl)	1724.82927	98	-0.005
(K)LGEYGFQNALIVR(Y)	42	N8(Deamidated)	1480.77873	98	0.004
(K)LGEYGFQNALIVR(Y)	13		1479.79204	97	0.000
(K)TVMENFVAFVVDK(C)	9		1399.69121	93	-0.001
(K)TVMENFVAFVVDK(C)	1	M3(Oxidation)	1415.68498	93	-0.003
(K)VPQVSTPTLVEVSR(S)	4		1511.84062	86	-0.002
(P)CTEDYLSLILNR(L)	3	C1(Carbamidomethyl)	1496.73870	83	-0.002
(R)RHPEYAVSVLLR(L)	6		1439.81028	82	-0.002
(M)ENFVAFVVDK(C)	7		1068.53703	77	0.001
(R)KVPQVSTPTLVEVSR(S)	2		1639.93723	76	-0.001
(K)DAFLGSFLYEYSR(R)	2		1567.73613	76	-0.007
(E)YGFQNALIVR(Y)	6	N5(Deamidated)	1181.63017	75	-0.001
(G)EYGFQNALIVR(Y)	5	N6(Deamidated)	1310.67229	74	0.000
(M)ENFVAFVVDK(C)	2	N2(Deamidated)	1069.51872	71	-0.001
(K)LGEYGFQNAL(I)	2		1111.54180	68	0.000
(K)KQTALVELLK(H)	13		1142.71416	68	-0.000
(Y)FYAPELLEYANK(Y)	1		1491.75029	68	-0.002
(K)LVNELTEFAK(T)	1	N3(Deamidated)	1164.61382	65	-0.001
<u>(E)DYLSLILNR(L)</u>	4		1106.62212	63	0.002
(K)DAIPENLPPLTADFAEDK(D)	2		1955.95671	60	-0.003
(S)TPTLVEVSR(S)	1		1001.56133	57	-0.001
(K)HLVDEPQNLIK(Q)	2		1305.71526	57	-0.001
(R)HPEYAVSVLLR(L)	2		1283.70691	56	-0.004
(P)EYAVSVLLR(L)	3		1049.59648	56	-0.003
(E)YAPELLEYANK(Y)	1		1344.68120	55	-0.002
(Q)VSTPTLVEVSR(S)	1		1187.66106	55	-0.002
(F)LGSFLYEYSR(R)	3		1234.60918	55	-0.001
(N)FVAFVVDK(C)	2		825.44951	55	-0.001
(N)LPPLTADFA(E)	7		944.50633	54	-0.001
<u>(A)IPENLPPLTA(D)</u>	5		1064.59661	54	-0.002
(Y)APELLEYANK(Y)	1		1181.61870	54	-0.001
(T)ALVELLK(H)	1		785.51219	54	-0.001
(E)YAVSVLLR(L)	2		920.55394	52	-0.002
(K)LVNELTEFAK(T)	1		1163.62810	52	-0.003
(Q)TALVELLK(H)	3		886.55974	48	-0.001
(G)FQNALIVR(Y)	1		960.56017	47	-0.002

Table 2. Proteome Discoverer results of a nanoRP-LC-ESI-OT-MS² (HCD) measurement after sequential digestion of BSA with trypsin and flavastacin. MASCOT search against unspecific *in-silico* digestion of mammalian taxonomy (UniProt-KB/SwissProt database). Peptide sequences with the N-terminal cleavage of aspartic acid or deamidated asparagine (as indicated from the supplier of flavastacin) are underlined and bold.

N₄₉₇ are described to carry complex-type mostly core-fucosylated and sialylated N-glycans, also including N-acetyllactosamine (LacNAc) extensions¹⁹. For site N₆₄₂ no N-glycosylation has been identified so far. The theoretical tryptic N-glycopeptides of hLTF are (R)PFLN₁₅₆WTGPPEPIEA AAVAR(F) (1964.0155 Da) (with a commonly missed cleavage due to proline: (R)TAGWNVPIGTLR PFLN₁₅₆WTGPPEPIEA AAVAR(F) (3229.7036 Da)), (R)TAGWNIPMGLLFN₄₉₇QTGSK(F) (2036.9812 Da) and (R)N₆₄₂GSDCPDK(F) (834.3178 Da). These peptides consist of 18/(30), 19 and 8 amino acids, respectively. The large peptide moieties of the sites N₁₅₆ and N₄₉₇ in combination with an N-glycosylation can reach masses, which are unsuitable for proper N-glycoproteomic analysis²⁰.

N-glycoproteomic analysis of flavastacin-generated hLTF N-glycopeptides revealed that all detected N-glycopeptides feature the N-glycosylated asparagine at the C-terminus, while the N-terminus was either a tryptic or an unspecific cleavage site. In addition, the generated N-glycopeptides were shorter compared to a solely tryptic digest. The identified N-glycopeptides for the N-glycosylation sites N₁₅₆ and N₄₉₇ are 4–16, and 4–13 amino acids long, respectively (see Table 1). In agreement with literature, no N-glycopeptide was detected for the N-glycosylation site N₆₄₂¹⁹. We did not perform any further enrichment of glycopeptides prior to MS analysis since the digestion strategy together with a long separation gradient and a high-resolution mass spectrometry measurement resulted in a comprehensive coverage of glycopeptides as shown in Fig. 2 (for solely tryptic digest see Supplementary Figure 2). The eluting N-glycopeptides are depicted by extracted ion chromatograms of MS/

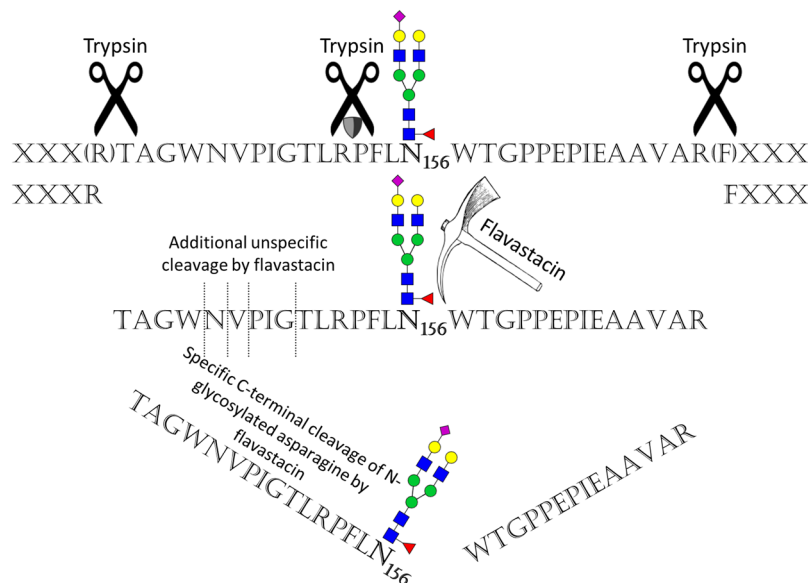


Figure 4. Specificity of flavastacin for the C-terminus of N-glycosylated asparagine. The amino acid sequence of the tryptic N-glycopeptide of hLTF with the N-glycan Hex₅HexNAc₄dHex₁NeuAc₁ linked to N-glycosylation site N₁₅₆ is shown. The scissors symbolize trypsin; the shield stands for cleavage inhibition of trypsin due to proline (P). Flavastacin is symbolized by a pick, and its specific C-terminal cleavage of N-glycosylated asparagine, as well as its unspecific N-terminal cleavages are shown. The N-glycan structure is illustrated according to CFG nomenclature²².

MS spectra containing glycan-specific oxonium ions: m/z 204.087 (N-acetylhexosamine (HexNAc) $[M + H]^+$), m/z 274.093 (N-acetylneuraminic acid (NeuAc)-H₂O $[M + H]^+$) and m/z 366.141 (hexose (Hex) + HexNAc $[M + H]^+$). The unspecific cleavage specificity due to flavastacin results in redundant N-glycopeptide signals over the elution time range with peptide moieties of different length. This can be a disadvantage for the analysis of more complex glycopeptide samples.

To illustrate the unique cleavage specificity of flavastacin, two examples for annotated MS/MS spectra of the N-glycopeptide containing sites N₄₉₇ (Fig. 3A) and N₁₅₆ (Fig. 3B) are shown. Both N-glycopeptides have the N-glycosylated asparagine at the C-terminus and the tryptic cleavage site at the N-terminus. For the site N₄₉₇, the fragment ion spectrum of the N-glycopeptide TAGWNIPM*GLLFN₄₉₇ with the N-linked glycan Hex₅HexNAc₄dHex₁NeuAc₁ (deoxyhexose (dHex)) (precursor ion: m/z 1184.1659 $[M + 3H]^{3+}$) is depicted, along with the corresponding peptide ion 1385.7266 $[M + H]^+$ (see Fig. 3A). The example for the fragment ion spectrum of site N₄₉₇ shows a neutral loss of 105 Da, related to a carbamidomethyl-methionine residue (see Fig. 3A). This neutral loss from a carbamidomethylated methionine has been described only rarely in literature²¹. Without the awareness of a carbamidomethylation of methionine, only a neutral loss of 48 Da would be recognized, which can mistakenly also be interpreted as a side chain loss of methionine sulfoxide. However, due to the specific digestion strategy involving both trypsin and flavastacin, in combination with high-resolution LC-MS, the identification of such unlikely modifications is also possible. For site N₁₅₆, the fragment ion spectrum of the N-glycopeptide TAGWNVPIGTLRPFLN₁₅₆ with the N-linked glycan Hex₅HexNAc₄dHex₂NeuAc₁ (precursor ion: m/z 1321.2479 $[M + 3H]^{3+}$) is depicted (see Fig. 3B) as well as the corresponding peptide ion 1755.9515 $[M + H]^+$. Both fragment ion spectra are dominated by the b-ion series, oxonium ions (B-ions) and the aforementioned fragmentation pattern: $[\text{peptide} - \text{NH}_3 + \text{H}^+]$; $[\text{peptide} + \text{H}^+]$, and $[\text{peptide} + \text{GlcNAc} + \text{H}^+]$ (Fig. 3). MS/MS spectra of additional N-terminal unspecifically cleaved N-glycopeptides are shown in Supplementary Figures 3 and 4.

Analysis of flavastacin-generated BSA peptides revealed primarily tryptic cleavage at the C-terminus (34 of 37 peptides), and tryptic and unspecific cleavage at the N-terminus (13 tryptic, 24 unspecific cleavages). Twelve peptides had tryptic cleavages at both termini (see Table 2). For the non-glycosylated BSA no flavastacin-generated peptides were detected with an asparagine at the C-terminus, which correlates to observations we made for non-glycosylated asparagines of the N-glycosylated hLTF (see Supplementary Table 1). To check for possible unspecific cleavages of the tryptic digest and its influence on the flavastacin digest, identified peptides of solely tryptically digested BSA (see Supplementary Table 2) and hLTF (see Supplementary Table 3) were examined. Here, almost exclusively, specific cleavages were identified in the tryptic digests of BSA and hLTF. This strongly suggests that the observed N-glyco-specific cleavage of hLTF, as well as the unspecific cleavage of BSA and hLTF of the combined digest (trypsin and flavastacin) can only be linked to the activity of flavastacin.

Whilst it has been described that flavastacin has specificities towards the N-terminus of aspartic acid, glutamic acid and cysteine, we found a unique cleavage specificity of flavastacin for the C-terminus of N-glycosylated asparagine, which was not explored up to now. All manually annotated hLTF N-glycopeptide related peptide sequences are listed in Table 1 (other identified non-glycosylated peptides are listed in Supplementary Table 1). Every single hLTF N-glycopeptide sequence has been cleaved at the N-glycosylated asparagine at the C-terminus

– independent of the N-glycoform attached to the respective N-glycosylation site. In addition, we observed that the N-terminus is a tryptic or an unspecific cleavage site. Based on the manually annotated peptide sequences and the database-assisted MASCOT search, no strict N-terminal cleavage of aspartic acid could be observed, neither for BSA nor for hLTF (Tables 1 and 2, Supplementary Table 1).

Conclusion and Outlook

Flavastacin shows a clear specificity for the C-terminus of N-glycosylated asparagine N₁₅₆ and N₄₉₇ in hLTF (illustrated in Fig. 4). Due to the presence of multiple N-glycosylation sites and the well-described complex-type N-glycan structures, hLTF is a very suitable glycoprotein to demonstrate this newly found specificity of flavastacin.

In contrast to previous work^{15,16}, we could not verify the claimed specificity of flavastacin for the N-terminus of aspartic acid, neither for hLTF nor for BSA. However, we could demonstrate that the sequential combination of trypsin and flavastacin for protein digestion successfully cleaves the N-glycoprotein hLTF in well-annotatable N-glycopeptide sequences. Interestingly, in contrast to unspecific digestion strategies using proteinase K or pronase, flavastacin works as an “N-glyco-specific” proteolytic enzyme (specific for N-glycosylated asparagine at the C-terminus). This property improves data quality as well as data analysis and therefore facilitates N-glycoproteomics significantly. However, the unspecific cleavage due to flavastacin at the N-terminus results in the distribution of redundant N-glycopeptide signals with peptide moieties of different length.

Overall, this finding improves the glycoproteomic toolbox and helps to overcome common problems in N-glycoproteomics, i.e. the presence of too large N-glycopeptides with too many amino acids and/or too many N-glycosylation sites for proper LC-MS analysis. Despite the fact that this specificity of flavastacin and its cleaving mechanism need to be examined also for other glycoproteins (as well as for more complex (glyco-) protein mixtures), and in particular for other types of glycosylation (like high-mannose-type, hybrid-type and O-glycosylation), the use of flavastacin will already be beneficial for glycoscience now, as it allows researchers to dig faster and deeper into N-glycoproteomes.

References

- Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
- Christiansen, M. N. *et al.* Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525–546, doi:<https://doi.org/10.1002/pmic.201300387> (2014).
- Rhodes, J., Campbell, B. J. & Yu, L.-G. Glycosylation and Disease. *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd: Chichester, doi:<https://doi.org/10.1002/9780470015902.a0002151.pub2> (2010).
- Jaeken, J. Congenital disorders of glycosylation (CDG): it's (nearly) all in it! *J Inher Metab Dis* **34**, 853–858, doi:<https://doi.org/10.1007/s10545-011-9299-3> (2011).
- Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448–462, doi:<https://doi.org/10.1038/nrm3383> (2012).
- Thaysen-Andersen, M. & Packer, N. H. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochimica et biophysica acta* **1844**, 1437–1452, doi:<https://doi.org/10.1016/j.bbapap.2014.05.002> (2014).
- Domon, B. & Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate journal* **5**, 397–409 (1988).
- Biemann, K. Appendix 5. *Nomenclature for peptide fragment ions (positive ions)*. *Methods in enzymology* **193**, 886–887 (1990).
- Scott, N. E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Molecular & cellular proteomics: MCP* **10**, M000031–MCP000201, doi:<https://doi.org/10.1074/mcp.M000031-MCP201> (2011).
- Hinneburg, H. *et al.* The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *Journal of the American Society for Mass Spectrometry* **27**, 507–519, doi:<https://doi.org/10.1007/s13361-015-1308-6> (2016).
- Drapeau, G. R. Substrate specificity of a proteolytic enzyme isolated from a mutant of *Pseudomonas fragi*. *The Journal of biological chemistry* **255**, 839–840 (1980).
- Tetaz, T., Morrison, J. R., Andreou, J. & Fidge, N. H. Relaxed specificity of endoproteinase Asp-N: this enzyme cleaves at peptide bonds N-terminal to glutamate as well as aspartate and cysteine acid residues. *Biochem Int* **22**, 561–566 (1990).
- Ingrosso, D., Fowler, A. V., Bleibaum, J. & Clarke, S. Specificity of endoproteinase Asp-N (*Pseudomonas fragi*): cleavage at glutamyl residues in two proteins. *Biochemical and biophysical research communications* **162**, 1528–1534 (1989).
- Kanie, Y. *et al.* Insight into the regulation of glycan synthesis in *Drosophila* chaoptin based on mass spectrometry. *PLoS One* **4**, e5434, doi:<https://doi.org/10.1371/journal.pone.0005434> (2009).
- Tarentino, A. L., Quinones, G., Grimwood, B. G., Hauer, C. R. & Plummer, T. H. Jr. Molecular cloning and sequence analysis of flavastacin: an O-glycosylated prokaryotic zinc metalloendopeptidase. *Arch Biochem Biophys* **319**, 281–285, doi:<https://doi.org/10.1006/abbi.1995.1293> (1995).
- Grimwood, B. G., Plummer, T. H. Jr. & Tarentino, A. L. Purification and characterization of a neutral zinc endopeptidase secreted by *Flavobacterium meningosepticum*. *Arch Biochem Biophys* **311**, 127–132, doi:<https://doi.org/10.1006/abbi.1994.1217> (1994).
- Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature methods* **6**, 359–362, doi:<https://doi.org/10.1038/nmeth.1322> (2009).
- Garcia-Montoya, I. A., Cendon, T. S., Arevalo-Gallegos, S. & Rascon-Cruz, Q. Lactoferrin a multiple bioactive protein: an overview. *Biochimica et biophysica acta* **1820**, 226–236, doi:<https://doi.org/10.1016/j.bbagen.2011.06.018> (2012).
- Yu, T. *et al.* Comprehensive characterization of the site-specific N-glycosylation of wild-type and recombinant human lactoferrin expressed in the milk of transgenic cloned cattle. *Glycobiology* **21**, 206–224, doi:<https://doi.org/10.1093/glycob/cwq151> (2011).
- Kolarich, D., Jensen, P. H., Altmann, F. & Packer, N. H. Determination of site-specific glycan heterogeneity on glycoproteins. *Nature protocols* **7**, 1285–1298, doi:<https://doi.org/10.1038/nprot.2012.062> (2012).
- Lapko, V. N., Smith, D. L. & Smith, J. B. Identification of an artifact in the mass spectrometry of proteins derivatized with iodoacetamide. *Journal of mass spectrometry: JMS* **35**, 572–575, doi:[https://doi.org/10.1002/\(SICI\)1096-9888\(200004\)35:4<572::AID-JMS971>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-9888(200004)35:4<572::AID-JMS971>3.0.CO;2-2) (2000).
- Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans†. *Journal of proteome research* **7**, 1650–1659 (2008).

Acknowledgements

T.N.K. and E.R. acknowledge support by European Union (EC) under the project “HTP-GlycoMet” (grant number 324400).

Author Contributions

A.P., M.H. and E.R. wrote the manuscript. A.P. did the measurements, analyzed the data and prepared all figures and tables. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-11668-1](https://doi.org/10.1038/s41598-017-11668-1)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017