Top-Down Sequencing of O-Glycoproteins by In-Source Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry for Glycosylation Site Analysis

ARTICLE *in* ANALYTICAL CHEMISTRY · JUNE 2011

Impact Factor: 5.64 · DOI: 10.1021/ac200493c · Source: PubMed

1 AUTHOR:

Franz-Georg Hanisch
University of Cologne
190 PUBLICATIONS **6,029** CITATIONS

SEE PROFILE

analytical chemistry

ARTICLE

pubs.acs.org/ac

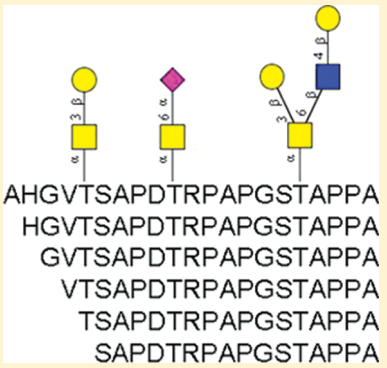
Top-Down Sequencing of O-Glycoproteins by In-Source Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry for Glycosylation Site Analysis

Franz-Georg Hanisch*

Institute of Biochemistry II, Medical Faculty, Joseph-Stelzmann-Strasse 52, and Center for Molecular Medicine Cologne, University of Cologne, Robert-Koch-Strasse 21, 50931 Köln, Germany

Supporting Information

ABSTRACT: The sites of mucin-type O-glycosylation are largely unpredictable, making structural analysis by mass spectrometry (MS) indispensible. On the peptide level, a site localization and characterization of O-linked glycans in situ using tandem MS with electron-transfer dissociation or matrix-assisted laser desorption ionization (MALDI) MS with postsource decay have been reported. The top-down sequencing on the protein level by MALDI-MS is based on the in-source decay (ISD) of intact glycoproteins induced by hydrogen radical transfer from the matrix. It allows a ladder sequencing from both termini with assignment of O-glycosylation sites based on intense c-, y-, and z-type ions. The feasibility of ISD-MALDI-MS in the localization of O-glycosylation sites was demonstrated with synthetic O-glycopeptides, the tandem repeat domain of recombinant MUC1, and the natural bovine glycoproteins asialofetuin and desialylated κ-casein. Ladder sequencing of the 17−18.5 kD MUC1 hexarepeat domains revealed (1) cell-specific glycosylation site patterns on comparison of probes expressed in human HEK-293 or Drosophila S2 cells, and (2) a site-specific microheterogeneity at the Thr/Ser sites with variations of the glycan compositions from zero to four monosaccharides. Novel O-glycosylation sites in the C-terminal domains of fetuin (T334) and κ-caseinoglycopeptide (S154 and T156) were assigned, the former representing a sequence conflict with the published T154.



The identification of O-glycosylation sites in endogenous or recombinant proteins is still challenging and lacking for most O-glycoproteins. Differing from N-glycosylation, where a consensus sequence controls the enzymatic transfer of a preassembled glycan to the protein backbone, the initiation of O-glycosylation by addition of *N*-acetylgalactosamine (GalNAc) to serine or threonine side chains is catalyzed by up to 20 isoenzymes with only partially overlapping substrate preferences and without a common sequon. This makes O-glycosylation sites largely unpredictable, although efforts in this direction were made by the creation of a software tool that is based on available structural data and neural network algorithms.[1]

Attempts to localize O-glycosylation sites in peptides by postsource decay (PSD) MALDI-MS were reported in 1997.[2] The process of metastable decomposition of molecular ions during their acceleration or when passing the drift region (postsource decay) is laser-induced and hence dependent on laser power and the density of the matrix cloud formed after the laser shot. As it is a soft fragmentation mode, PSD shows some degree of selectivity with respect to peptide bond cleavage, making it more dependent on the intrinsic chemical properties of the analyte compared to collision-induced dissociation (CID)-induced fragmentation. On the other hand, it offers the advantage that labile-bound O-glycans remain stable during the fragmentation process. Initial attempts to localize O-glycosylation sites in glycopeptides by ESI mass spectrometry were undertaken on instruments with a q-tof configuration of mass analyzers and CID for fragment generation.[3] A technical limitation of CID for site-specific O-glycopeptide analysis is the preferential fragmentation of the labile glycosidic bonds. Accordingly, the glycan is lost by fragmentation before the peptide backbone breaks and attachment sites cannot be localized by the respective mass incremental increases of Ser or Thr residues. Some ESI instruments with ion traps or with an orbitrap offer electron-transfer-dissociation (ETD) as an alternative fragmentation mode.[4] ETD is based on the transfer of electrons from a donor compound (like fluoranthen) to the analyte. As the transfer occurs in a random fashion and the radical formed induces breakage of the respective bond, there is no preferential cleavage of labile bonds as observed in CID, and fragment intensity shows a more even distribution over the entire peptide. The consequence is that longer stretches of the peptide sequence can be read on the basis of continuous c- or z-ion series and that labile modifications can be localized. However, recent evidence revealed that glycopeptides with $m/z > 850$ frequently feature precursor ions of low charge density that will not undergo efficient ETD fragmentation.[5]

Received: February 25, 2011
Accepted: April 28, 2011
Published: April 28, 2011

In-source decay (ISD) MALDI spectra are known to provide sequence information directly from highly purified intact proteins and can be used for partial top-down N- and C-terminal sequencing of proteins. ISD-MALDI mass spectrometry was introduced in 1995 by Brown and Lennon[6] and was later applied in the reflectron technology by Suckau and Cornett.[7] During ISD-MALDI-MS acquisitions, the molecular ions fragment spontanously within the ion source and yield preferentially c-type ions from the N-terminus and y- or z-type ions from the C-terminus of the protein. The type of fragment ions formed in the MALDI process was claimed to depend to some extent on the matrix used in sample preparation. While dihydroxybenzoic acid (DHB), and in particular its mixture with an isomer called "super-DHB" (sDHB), was found to give rise to fragmentation of both the N- and C-terminus of proteins by ion formation of the c- and y- (or z-) series, 1,5-diaminonaphthalene (DAN) was demonstrated to yield largely c-type ion series.[8] Technical developments in MALDI-TOF instrumental features[9] and in sample preparation[10] have led to significant improvements in the performance of ISD-based sequencing of proteins. Currently, spectra can be recorded with mass accuracies in the 10 ppm range, and reading lengths of peptide sequences up to 70 aa residues can be achieved. Because of a mass cutoff (ions <1.000 Da have to be deflected), ions can be observed typically between 1.000 and 8.000 Da, provided that the sequence contains no disulfide bridges within this region. Another limitation is found in short gaps in the c-ion series caused by proline residues. To extend sequence information toward the N- and C-terminus of the protein, ladder fragments from in-source decay can be selected as precursor ions for post-source decay (PSD) fragmentation based on laser-induced dissociation (LID).[8] Compared to Edman sequencing, the MS-based technique offers advantages by being less expensive and time-consuming; also, C-terminal sequences can be read and proteins with N-terminal modifications are readily sequenced.[8] However, top-down sequencing (TDS) by ISD-MALDI suffers from low sensitivity and the restriction to pure protein samples.

Post-translational protein modifications, even labile bound glycans linked O-glycosidically to Ser/Thr, remain bound to the protein during ladder fragmentation in the ion source (refer to the present study). This is due to the fact that fragmentation is induced by hydrogen radical transfer from the matrix, a process related to ETD. No energy relocation to labile bonds, as in CID, occurs on the subnanosecond time scale. The soft fragmentation mode with stochastic, nearly equi-intense ion formation should hence be well suited for the site-specific analysis of modified proteins.

The current investigation represents the first comprehensive evaluation of ISD protein fragmentation for the N- and C-terminal sequencing of O-glycoproteins and for the localization of O-linked glycans by ladder sequencing. A great advantage for unequivocal assignments of modified sites is the splitting of c- or y-ion series into multiple branches caused by site-specific glycosylation microheterogeneity, which adds to the overall confidence of the assignment. The fragments are formed under conditions that leave O-glycosidic bonds largely unaffected. Through application of this technique, peripheral stretches of MUC1 tandem repeat domains became accessible to ladder sequencing and in situ localization of complex glycans. One novel O-glycosylation site was assigned to bovine fetuin, which is in accordance with findings on the glycopeptide level after partial deglycosylation.[5] Moreover, one sequence ambiguity became apparent and two novel glycosylation sites were localized within the C-terminal glycopeptide region of bovine κ-casein.

EXPERIMENTAL SECTION

Materials. Glycoproteins: Bovine asialofetuin and κ-caseine were purchased from Sigma (Steinheim, Germany).

Glycopeptides: A test glycopeptide was synthesized that corresponds to two repeat units of the variable number of tandem repeats domain of MUC1 and contains GalNAc residues at Thr9 and Ser29: HGVTSAPD**T**RPAPGSTAPPAHGVTSAPE**S**RPAPGSTAPPA (Biosyntan, Berlin, Germany). Glycopeptides H11 to H13 were kindly provided by Prof. Hans Paulsen, University of Hamburg, Germany. Matrixes: 1,5-diaminonapthalene (DAN) was purchased from ACROS Organics (Belgium) and used without further purification; dihydroxybenzoic acid was from Bruker Daltonics (Bremen, Germany). Solvents: acetonitrile (HPLC gradient grade, Sigma, Steinheim, Germany); water (for HPLC) and water with 0.1% TFA (Fluka, Buchs, Switzerland); formic acid 50% (Fluka).

Generation and recombinant expression of MUC1-VH in *D. melanogaster* S2 cells, its affinity isolation by Ni chelate chromatography, and purification on a C8 reversed-phase column were described previously.[11] The generation of MUC1-S,[12] the recombinant expression in human HEK-293 cells, and its isolation[13] were previously described. An amount of 10 μg of the fusion proteins was digested with endoproteinase LysC (1 μg/in 20 μL ammonium hydrogencarbonate, pH 8.0) for 4 h at 37 °C. The dried products were taken up in 100 μL of water and chromatographed on 0.48 × 250 mm C8 reversed-phase column (vydac) in a gradient from 8 to 80% acetonitrile/0.1%TFA during 30 min. Fractions of the eluate were analyzed by MALDI-TOF mass spectrometry in the reflectron mode to identify the LysC cleavage products. The hexarepeat domain TR6, which is identical with respect to the amino acid sequence in both constructs, was also analyzed in the linear mode to determine the exact molecular mass of the glycopeptides.

Top-down Protein Sequencing by in-Source Decay for Site Localization of O-Linked Glycans. *Sample Preparation.* For sample preparation, 100 pmol/μL of the glycopeptide or glycoprotein was solubilized in water/0.1% TFA. For matrix preparation, a saturated colorless to pale violet solution of DAN was prepared in 50% acetonitrile/0.1% aq TFA. The sample was applied onto the target by mixing 0.5 μL of sample with 1.0 μL of matrix prior to application of 1.5 μL onto the stainless steel MALDI target. The sample spots were dried at ambient temperature, and the quality of the DAN matrix preparation was inspected on the video image to ensure homogeneous needle-like crystallization. If crystallization was heterogeneous, another aliquot of pure matrix solution was added to the spot for recrystallization. Alternative matrixes, such as DHB and sDHB, were used by applying the dried-droplet method (equal volumes of saturated solutions of the respective matrixes in 33% acetonitrile/0.1% aq TFA).

Mass Spectrometry. ISD-MALDI-MS analysis was performed on an UltrafleXtreme MALDI-TOF-TOF mass spectrometer (Bruker Daltonics) using the acquisition software FlexControl 3.3 and the data evaluation software FlexAnalysis 3.3 (Bruker Daltonics). The following parameters were used: Ion source 1 (25.00 kV); ion source 2 (22.25 kV); lens (7.50 kV); reflector 1 (26.50 kV); reflector 2 (13.45 kV); pulsed ion extraction (80 ns); matrix suppression by deflection (900 Da). The mass range selected for detection was from 1.000 to 8.000 Da. In general, 5.000−10.000 laser shots were accumulated at 1 kHz acquisition for each sample spot.

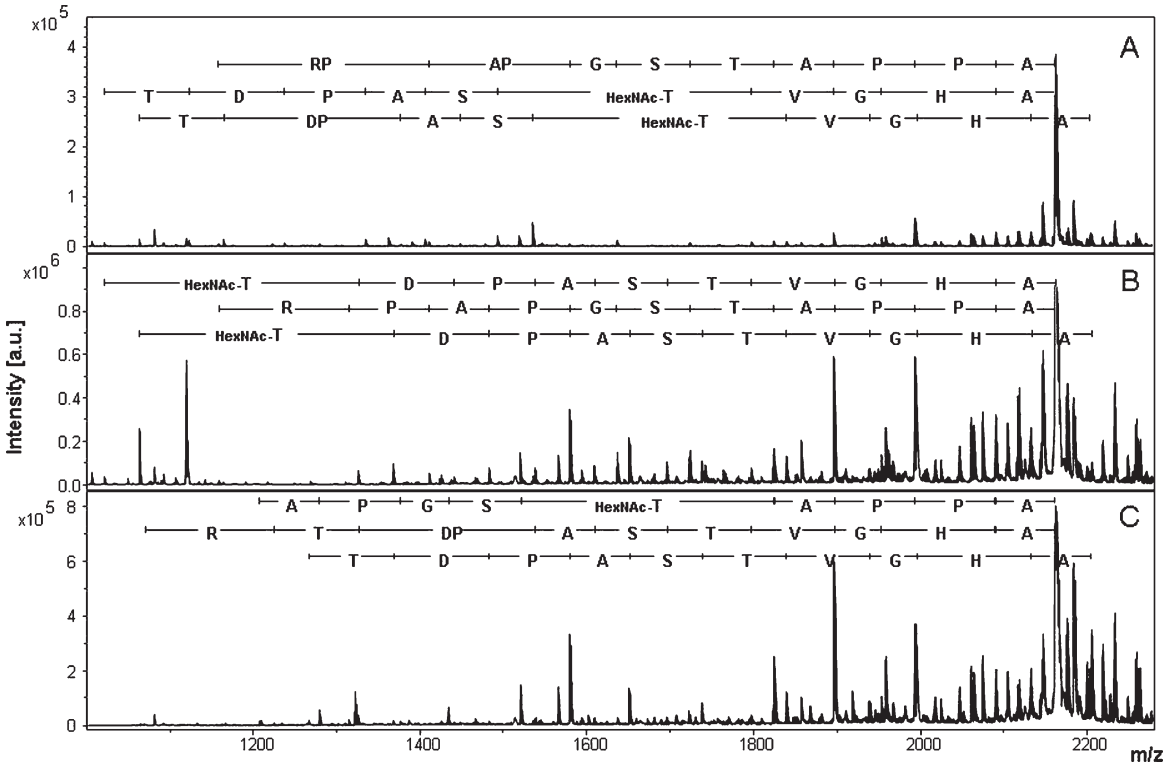


Figure 1. Determination of O-glycosylation sites in 21-meric MUC1 glycopeptides. A, glycopeptide H11 carries GalNAc at position Thr5; B, glycopeptide H12 carries GalNAc at Thr10; C, glycopeptide H13 carries GalNAc at Thr17. H11 to H13 cover one tanden repeat of human MUC1 plus one amino acid (AHGVTSAPDTRPAPGSTAPPA). Synthetic O-glycopeptides were applied in DHB matrix and measured by ISD-MALDI-MS as detailed in Experimental Section.

Automatic processing by FlexAnalysis was chosen using the SNAP algorithm for monoisotopic peak annotation and internal calibration. During this procedure the reflector ISD spectrum of the mass calibrant (BSA) is internally calibrated using a mass control list, which contains theoretical c-ion masses from 1192.62 to 5784.6 Da. This calibration is automatically applied to the sample spectra. Spectra were analyzed in FlexAnalysis (Bruker Daltonics) using the manual annotation tool and predefined mass incremental building blocks for amino acids, X-Pro diads, and glycosylated Ser/Thr to determine sequence stretches and glycosylation sites. Evaluation of spectra can be supported by automatic annotation using BioTools 3.2 (Bruker Daltonics).

N-Terminal Fixed-Charge Tagging. To 1 μL of peptide/glycopeptide in aqueous solution (100 pmol/μL) was added 2 μL of Tris-HCl buffer, pH 8.2, followed by 1 μL of 1 mM (*N*-succinimidyloxycarbonylmethyl)tris(2,4,6 trimethoxyphenyl)-phosphonium bromide (TMPP-AcOSu bromide) reagent in ACN/water (2:8 v/v). After 20 min incubation with sonication in a water bath, the reaction mixture was dried by vacuum rotation and taken up in 10 μL of 0.1% aq TFA prior to loading onto a ZipTipC18 tip. Elution of tagged (glyco)peptides was done by pipetting preapplied matrix solution (20 mg of DHB in 50% ACN/0.1% aq TFA) several times.

RESULTS

Determination of O-Glycosylation Sites in Synthetic O-Glycopeptides of Varying Lengths. To demonstrate the applicability of ISD ladder sequencing for shorter glycopeptides, five 21-meric O-glycopeptides were analyzed that correspond to one tandem repeat of human MUC1 plus one amino acid and are substituted with either GalNAc (M 2160.045) or Gal-GalNAc (M 2322.098) at varying positions of five potential glycosylation sites. Theoretically, the lower size limit for ISD ladder sequencing should be in the range of 20-mers. Although up to ten of the N- or C-terminal peptide ladder ions with masses below 1000 Da are deflected, the remaining c- or y/z-type ions of higher masses should cover the entire 20-meric peptide sequence and give unequivocal information on the glycosylation sites. Glycopeptide lengths below 20 aa are outside of the suitable size range. To enhance C-terminal fragmentation, the glycopeptides were applied in DHB or sDHB matrix. In compound H11 (Figure 1A, Table S-1A, Supporting Information) the HexNAc residue could be localized to threonine at position p5 (T5) as indicated by ions of the y-series at m/z 1493.2 and 1797.4 corresponding to a mass increment of HexNAc-T (304.2 Da). Assignment of the glycosylation site is indicated by absence of incremental mass shifts for T10 (y-ion series) and S16 and T17 (c ion series). Because of incomplete removal of O-acetylation from synthetic O-glycopeptides, molecular ions MH + 42 were detectable with low intensity in all standard compounds (for H11−H13 see Figure 1; A2M1 and A5, not shown). However, these contaminants yielded independent strong c- and y-type ion series (only the latter was indicated in spectra shown in Figure 1, leaving some major ions nonannotated) and supported glycosylation site assignments. Analogous to compound H11, the glycosylation site in H12 was clearly assigned to T10 as indicated by the c-ion at m/z 1157.9 and the y-type ions at m/z 1021.9 and 1326.1 (Figure 1B, Table S-1B). In the ISD spectrum of compound

dx.doi.org/10.1021/ac200493c |*Anal. Chem.* 2011, 83, 4829–4837

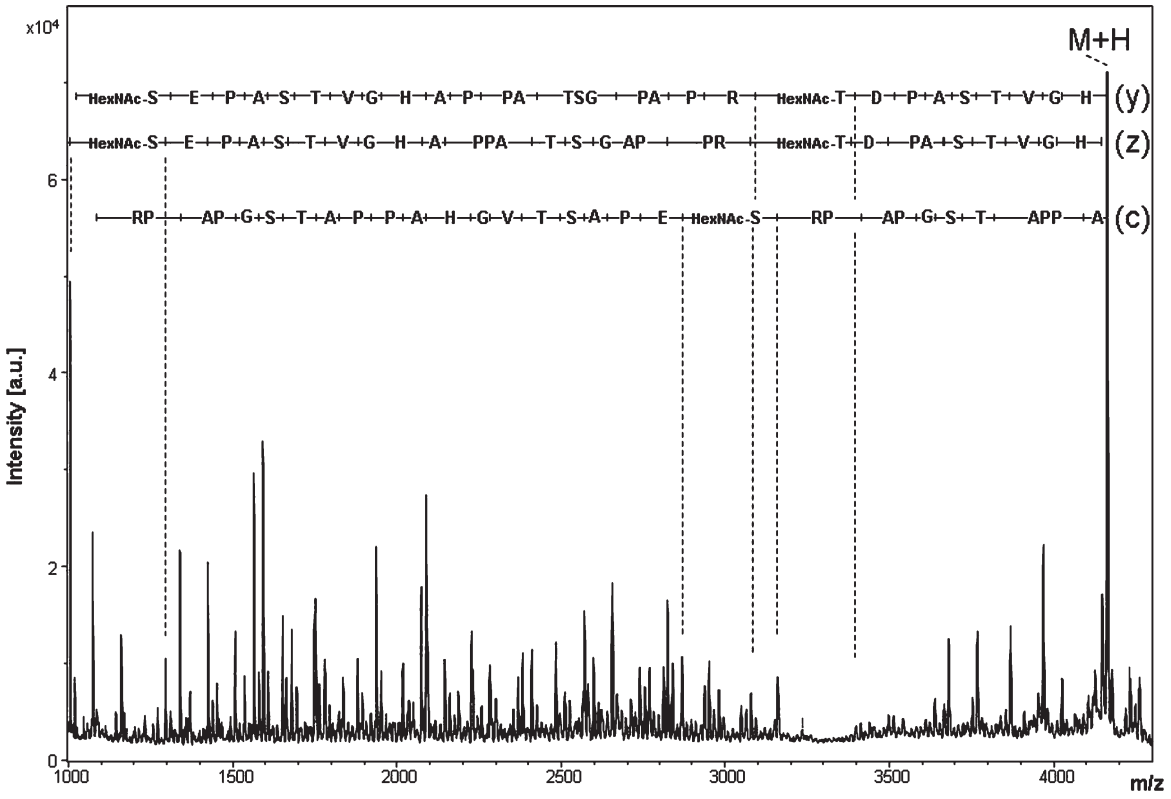


Figure 2. Determination of O-glycosylation sites in 40-meric MUC1 glycopeptide. The synthetic O-glycopeptide comprised two tandem repeats of human MUC1 starting at the HGVT motif (HGVTSAPDTRPAPGSTAPPAHGVTSAPESRPAPGSTAPPA). Positions 8/9 and 28/29 show a sequence polymorphism. Two GalNAc residues are linked to positions Thr9 and Thr29. The analyte was applied in DHB matrix and analyzed by ISD-MALDI-MS as described in Experimental Section.

H13 the glycosylation site at T17 was indicated by the respective c-ions at m/z 1521.3 and 1825.5 (Figure 1C, Table S-1C).

The sequence variation in compound A2M1 (isobaric replacement of D9-T10 by E9-S10) and the location of the Hex-HexNAc disaccharide (monoisotopic mass incremental shift of 365.13 Da) were indicated by the z+2 ions at m/z 1009.3, 1462.3 and 1591.5, respectively (not shown). In compound A5 the disaccharide was localized at S16 based on c-type ions at m/z 1435.3 and 1888.1 (not shown).

The synthetic 40-meric MUC1 peptide (calculated MH 4160.005 Da, experimentally determined MH 4161.05) covers two tandem repeats of the mucin and carries GalNAc residues at T9 and S29, respectively. On ISD ladder sequencing, three continuous ion series (c, y, z) were formed that allowed unambigous assignment of the two glycosylation sites and of the sequence variations in the peptide core (Figure 2, Table S-2, Supporting Information). In detail, the glycosylation of the E28-S29 motif became apparent by the ladder fragments at m/z 1005.6 and 1295.8 (z-type ions) or at m/z 1020.7 and 1310.7 (y-ions), whereas HexNAc substitution of the D8-T9 motif was indicated by ions at m/z 3077.9 and 3383.1 (z-type ions) or at m/z 3092.9 and 3397.1 (y-type ions). Within the c-type ion series, the signals at m/z 1086.5 and 3158.1 correspond to glycosylated T9 and S29.

O-Glycosylation Sites in MUC1 Tandem Repeat Domain. The fusion protein MUC1-VH comprises a truncated tandem repeat domain of six 20-meric repeats that is flanked by designed LysC cleavage sites and exhibits oligo-His and VH tags at the C-terminus (Figure 3, Table S-3, Supporting Information). The hexarepeat peptide TR6 (calculated mass of unglycosylated peptide: 12361 Da) was cut from the protein by LysC digestion, and the HPLC purified product contained in fraction 1 (Figure 3A, Table S-3A) was analyzed by MALDI-TOF mass spectrometry in the linear, positive ion mode for molecular mass determination of the 131-meric glycopeptide (Figure 3B, Table S-3B). The median of the peak cluster at m/z 17190 indicated that about 13 HexHexNAc disaccharides are linked to the hexarepeat peptide indicating two substituted sites per repeat. The Thomsen–Friedenreich (TF) disaccharide, Gal1-3GalNAc, was previously found to represent the predominant glycan linked to MUC1-VH.[11] By ISD top-down sequencing of the hexameric repeat peptide TR6, it was possible to assign 40 N-terminal and 30 C-terminal amino acid residues corresponding to over 50% of the entire fusion protein sequence. The fragment series of the c- and y-ion type generated in the DHB matrix revealed that exclusively the Thr residues in the VTSA and GSTA motifs were glycosylated (Figure 3C, Table S-3C). Some positions showed microheterogeneity by partial substitution with zero to two sugars. Six of the putative twelve Thr residues in VTSA and GSTA motifs were characterized as glycosylated, whereas all Ser residues of the accessible terminal repeats and the respective PDTR motifs were found to be unglycosylated. Expecting the same substitution pattern for the internal repeats a total number of 12 substituted sites would well agree with the calculated degree of glycosylation based on the total mass of TR6.

The human glycosylation probe MUC1-S expressed in human HEK-293 cells was digested in the same way, and the HPLC purified TR6 glycopeptide revealed a median molecular mass

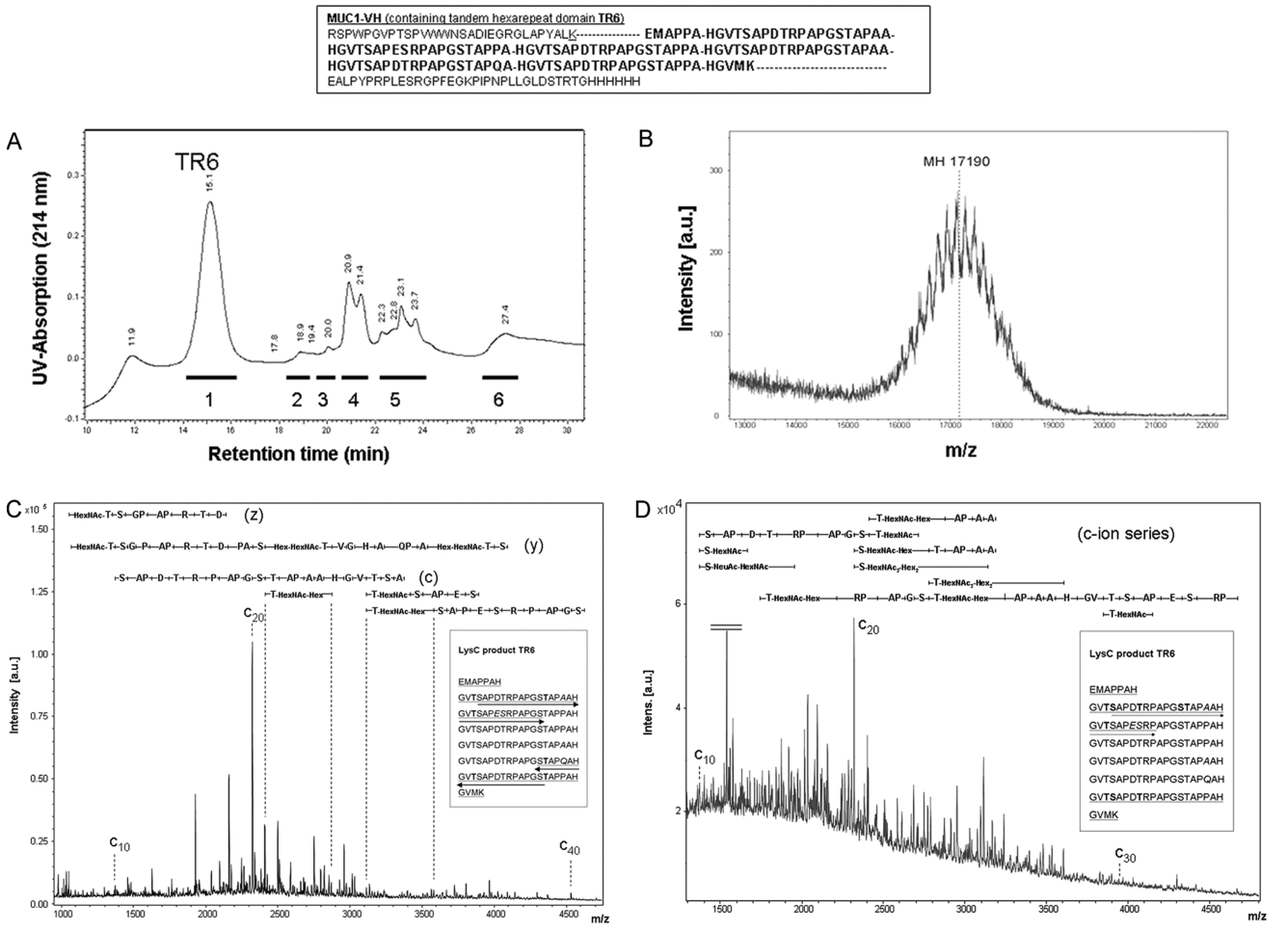


Figure 3. Top-down glycoprotein sequencing of a hexarepeat domain in recombinant hMUC1 expressed in *D. melanogaster* S2 cells or human HEK-293 cells. (A and B) At the top of the figure, the amino acid sequence of secretory fusion protein MUC1-VH (expressed in Drosophila S2 cells[11]) is shown in the single-letter-code. The sequence in boldface refers to the tandem hexarepeat domain TR6, which was excised by proteolysis with LysC. A, TR6 was separated from N- and C-terminal peptides by reversed-phase C8-HPLC (fraction 1). B, TR6 was analyzed by linear MALDI-TOF analysis to determine the molecular mass of the glycopeptide. C, Ladder sequencing of MUC1-VH-derived TR6 by ISD-MALDI-MS revealed strong c- and y-type ion series allowing unequivocal asignment of six O-glycosylation sites (refer to the insert: the sequenced N- and C-terminal peptide regions are underlined, the region covered by c-type ladder ions is marked by an arrow, sequence variations are given in italics, and O-glycosylated threonines are highlighted in boldface). D, ISD-MALDI-MS of TR6 from human HEK-293 cells revealed a much more complex N- and C-terminal fragmentation pattern. Only c-type ladder ions were considered for a sequencing of N-terminal streches and the localization of O-glycans (refer to the insert: the sequenced N- and C-terminal peptide regions are underlined, the region covered by c-type ladder ions is marked by an arrow, sequence variations are given in italics, and O-glycosylated threonines are highlighted in boldface).

near m/z 18000. This higher molecular mass of MUC1-S compared to MUC1-VH reflects the previously described higher density and complexity of O-glycosylation on the mammalian probe.[12,13] As shown for MUC1-VH, a similar reading length was accessible in ISD-MALDI-MS of MUC1-S-derived TR6 glycopeptide based on the c-type ion series (Figure 3D, Table S-3D). In contrast to the MUC1-TR6 from *D. melanogaster* cells, the human probe from HEK-293 cells exhibited glycan modifications at each of the five potential sites per repeat, and the glycan chains varied in size from one to six monosaccharides (refer also to Discussion).

O-Glycosylation Sites in Desialylated Bovine κ-Casein. The desialylated cow κ-casein was analyzed by top-down protein ladder sequencing using two alternative matrixes. The fragmentation in DAN matrix (not shown) revealed strong c-type ions (c11-c33) with unequivocal assignment of the N-terminal sequence CEKDERFFSDKIAKYIPIQYVLS (p32−54).

On the other hand, the spectrum registered for the sDHB sample (Figure 4, Table S-4, Supporting Information) was dominated by z/y-type ions from the C-terminus. Sequencing of the glycopeptide region (p142−190) was facilitated by initial cleavage of the C-terminal peptide p169−190 within the ion source that was followed by ladder fragmentation of the resulting truncation product. The experimental data (seven overlapping ion series diverging at heterogenously O-glycosylated Ser/Thr) covered the sequence stretch p150−159. However, the seven ion series were consistently in sequence conflict with respect to position p154, where T was replaced by S. Further discrepancies with previously reported experimental data became evident on comparison of the revealed glycosylation sites. While Pisano et al.[14] had identified T152, T154, and T157 as O-glycosylated, the ISD spectrum in Figure 4 indicated the five O-glycosylation sites T152, S153, S154, T156, and T157. The two novel O-glycosylation sites S154 and T156 were secured by

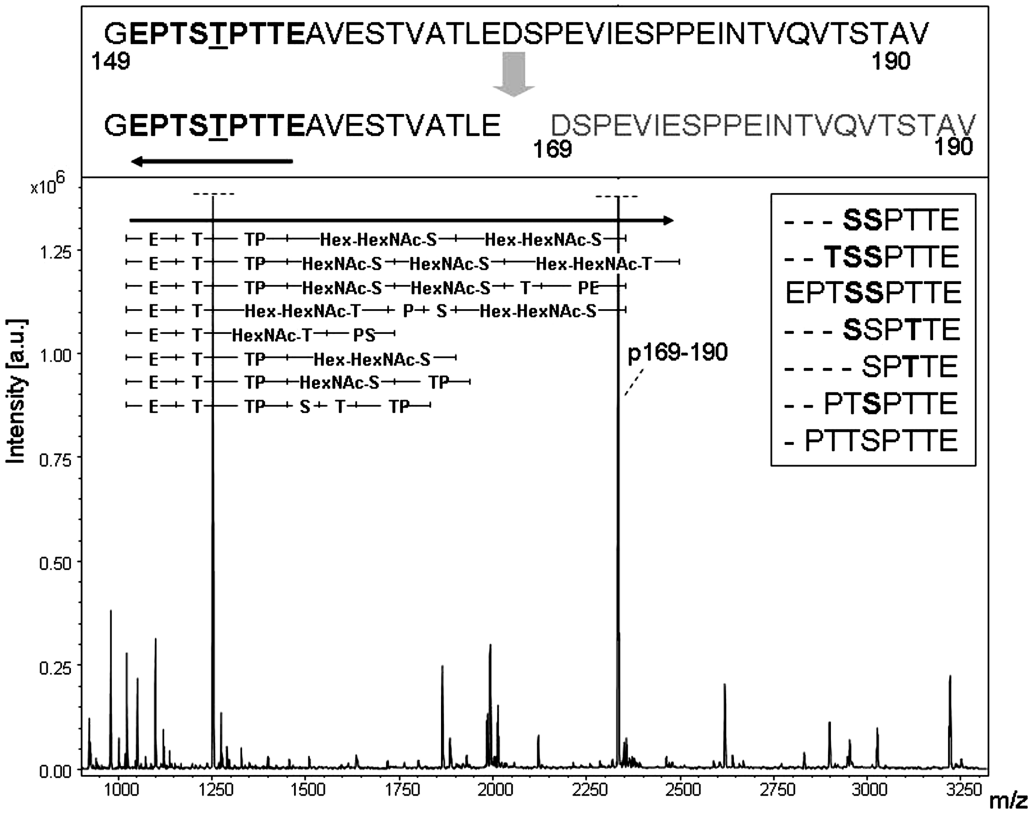


Figure 4. Top-down glycoprotein sequencing of bovine k-caseine. The sample was applied in sDHB matrix and analyzed by ISD-MALDI-MS. Sequences shown on top of the figure refer to the C-terminal domain of the protein, which spontanously fragments in the ion source to yield p169−190 and a correspondingly truncated κ-caseinoglycopeptide. The arrows indicate the reading direction of the amino acid sequence or of the z-ion series annotated in the mass spectrum. The insert shows the readable section of clustered Ser/Thr (p150−158). The gray-shaded serine indicates a sequence ambiguity with published data. Glycosylated Ser/Thr are highlighted in boldface.

fragment ions corresponding to the respective unglycosylated or the HexNAc and Hex-HexNAc substituted species.

O-Glycosylation Sites in Bovine Asialofetuin. As observed for κ-casein, the different matrix preparations of asialofetuin provided distinct sequence information on ISD-MALDI-MS. The ladder sequencing of the C-terminal portion of asialofetuin covered amino acid positions p321−352 of the protein, irrespective of the matrix type. The remaining C-terminal positions p353−359 were confirmed by TOF−TOF analysis of ladder fragment z10 registered with high intensity in the sDHB spectrum (not shown). However, N-terminal fragment ladders started at different positions in the peptide core. While for the DAN sample the sequence could be read from p30−50 (Figure 5A, Table S-5A, Supporting Information), the sDHB spectrum revealed initial fragmentation at Lys-C (p50) and ladder fragment formation of the truncated protein from p51−77 (Figure 5B, Table S-5B). The peptide regions of fetuin covered by ISD-MALDI-MS allowed the verification of one established O-glycosylation site S341, which was found in accordance with the published data to be either unsubstituted or substituted with Hex-HexNAc (Figure 5A and 5B).[15] Moreover, evidence was obtained for the O-glycosylation of T334, a site which has recently been found in another study to be at least substoichiometrically glycosylated.[5] At this site, either HexNAc (major) or Hex-HexNAc (minor) was detected upon analysis of samples in sDHB, whereas only Hex-HexNAc was found in the DAN samples. The heterogeneity at this glycosylation site is most likely not induced by the different matrixes, since the measurement of synthetic glycopeptides with Gal-GalNAc substitution (A2M1 and A5) never did reveal any matrix-dependent glycosylation heterogeneities.

Evaluation of N-Terminal Fixed-Charge Tagging for Increased Ion Intensities and Fragmentation. The preferential formation of singly charged ions in the MALDI process and the transfer of hydride radicals from matrix molecules to analyte ions during initiation of in-source decay result in comparatively low fragment yields and hence an overall low sensitivity of TDS. To address this limitation, the attempt was made to increase fragment ion yield of the c-ion series by fixed-charge tagging of the N-terminus. Low picomole quantities of peptide or protein were modified with (*N*-succinimidyloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-O-Su) to generate the respective N-terminal TMPP-Ac derivatives, while the ε-NH_2 groups of lysines remain unchanged.[16] By elution of modified or unmodified glycopeptides into the preapplied matrix and comparison of the respective ISD spectra, a considerable reduction in fragmentation efficiency was observed for the tagged species (data not shown). However, N-terminal tagging with TMPP-Ac may have the advantage that the readable sequence is shifted by approximately five amino acids toward the N-terminus.

DISCUSSION

In the past, efforts were made to localize alkali-labile bound O-glycosylation sites by substitution chemistry, the base-catalyzed

dx.doi.org/10.1021/ac200493c |*Anal. Chem.* 2011, 83, 4829–4837

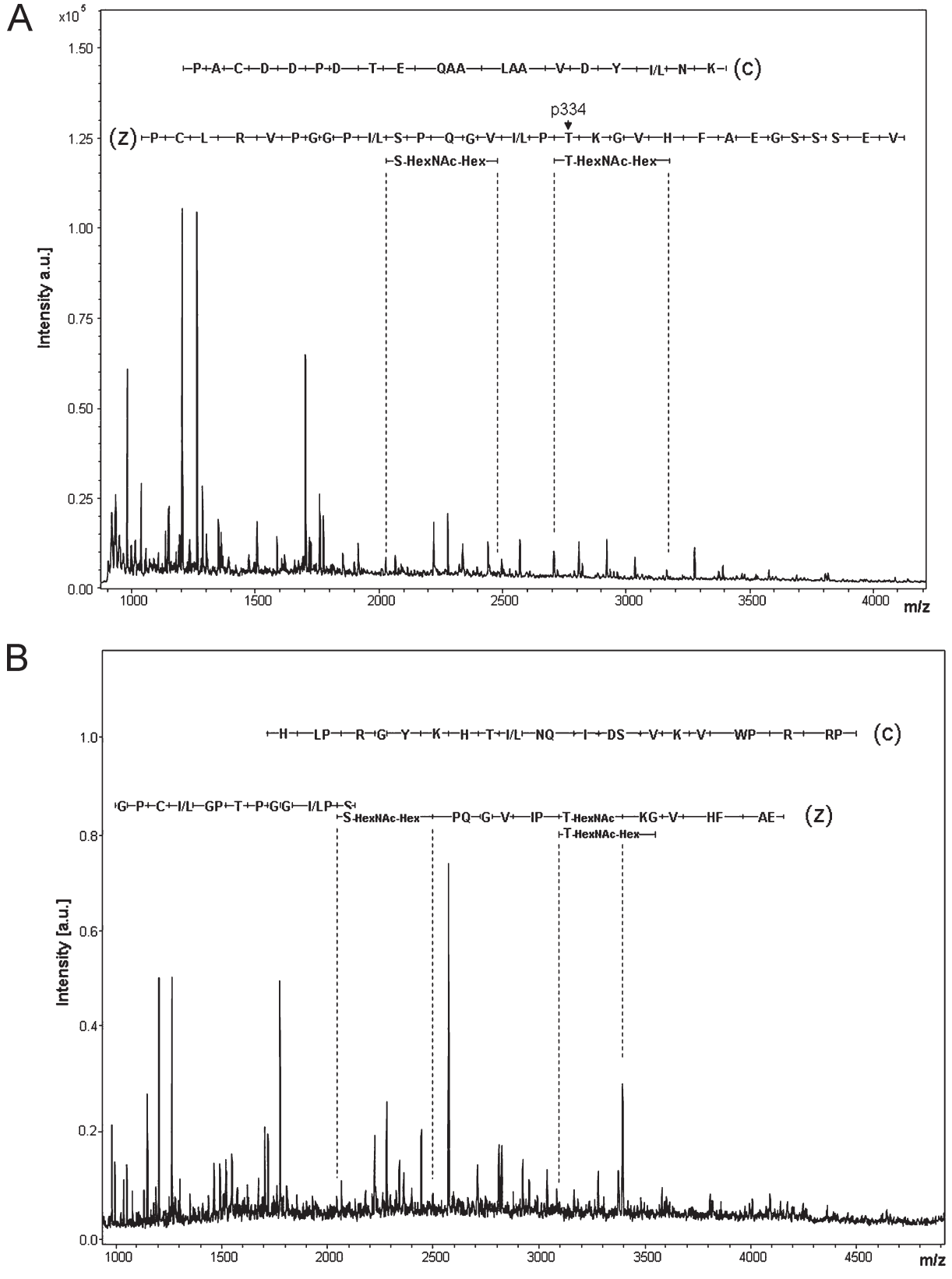


Figure 5. Top-down glycoprotein sequencing of bovine asialofetuin. A, The ISD spectrum shown refers to bovine asialofetuin applied in DAN matrix. Readable sequences cover the N-terminal peptide stretch p30−50 (ion series c30 − c50) and the C-terminal region p321−351 (ion series z9−z39). While position Ser341 has been known for decades to be substoichiometrically O-glycosylated, Thr334 was only recently described to carry O-linked glycans.[5]

dx.doi.org/10.1021/ac200493c |*Anal. Chem.* 2011, 83, 4829–4837

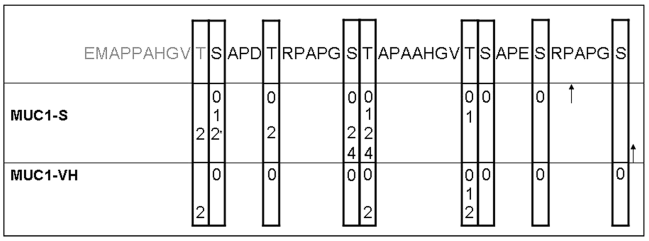


Figure 6. Summary of O-glycosylation patterns on N-terminal regions of the MUC1 tandem repeat domain. The LysC cleaved tandem repeat domain of recombinantly expressed MUC1-S (HEK-293 cells[12,13]) or MUC1-VH (*D. melanogaster* S2 cells[11]) was sequenced by ISD-MALDI-MS. A site-specific microheterogeneity of Thr/Ser (T/S) sites was revealed with HexNAc (1), Hex-HexNAc (2), NeuAc-HexNAc (2*), Hex2-HexNAc2 (4), or no glycan (0) in the N-terminal two repeats. Arrows indicate the readable sequence based on c-type ions.

β-elimination of glycans, and the addition of nucleophilic compounds to yield a labeled product with a defined mass tag. These so-called β-elimination−Michael addition (BEMAD) approaches using ammonia, various alkylamines, or DTT[17−21] show some advantages in MS analysis but suffer from the disadvantage that information on the glycan structure is lost. Other approaches were based on the partial acid hydrolysis of O-glycopeptides in the gas phase but were applicable only for small glycan chains lacking labile monosaccharide constituents, such as sialic acid or fucose.[22] With the advent of soft fragmentation modes in tandem mass spectrometry, the direct mass spectrometric analysis of native O-glycopeptides enables the elucidation of both the localization of glycan chains in the peptide backbone and the structural features of the glycan in situ. Besides ECD (electron-capture dissociation) and ETD fragmentation, the ISD in MALDI-MS could become a feasible and potent alternative for O-glycan site localization.

ISD analysis of intact glycoproteins may suffer from only partial access to more external protein sequences by being restricted to about 70 residues of the terminal regions. This restriction can become even stronger with increasing site heterogeneity introduced by partial glycosylation or glycan heterogeneity. However, the limitation can be overcome by enzymatic prefragmentation with endoproteinases, such as for example LysC, as shown for the recombinant fusion proteins MUC1-VH and MUC1-S. We could isolate the entire tandem repeat domain of the minimucins by specific proteolysis at designed LysC cleavage sites and sequence 50% of the 17−18.5 kD glycopeptides. About two N-terminal tandem repeats exhibiting variable amino acid sequences (DT vs ES and P vs A) were accessible to ladder sequencing (Figure 6) and revealed (1) cell-specific glycosylation site patterns on comparison of probes expressed in human HEK-293 or Drosophila S2 cells, which are in accordance with previous glycan analyses,[11,13] and (2) a site-specific microheterogeneity at the Thr/Ser sites with variations of the glycan compositions from zero to four monosaccharides.

In other cases (as shown for κ-caseine and fetuin), an in-source prefragmentation occurs, generating larger glycopeptides, which then fragment spontanously by further ISD to yield peptide ladders. This internal prefragmentation, however, is unpredictable and cannot be controlled by the investigator. A technical solution could be also the partial predigestion with exopeptidases, such as carboxypeptidase Y, followed by LC purification. Omission of LC purification of the respective glycopeptides from the complex mixture of digestion products would yield ISD spectra of overwhelming complexity due to accumulating isobaric or nearly isobaric fragment ions.

Taking these considerations into account, ISD-MALDI top-down sequencing exhibits no obvious weaknesses compared to PSD-MALDI (or ETD-ESI) bottom-up sequencing of proteolytically derived glycopeptides. Both analytical approaches differ with respect to the mode of fragmentation and the degree at which labile post-translational modifications are retained. While in PSD the metastable decomposition of molecular ions is laser induced and hence dependent on the laser power and intrinsic molecular properties of the analyte, ISD resembles the ETD process of peptide fragmentation and is the softer, more stochastic fragmentation mode with no apparent loss of sugar. On the contrary, PSD-MALDI suffers from partial loss of sugar, lower resolved TOF-TOF spectra, and longer gaps resulting from the more selective peptide fragmentation. Irrespective of this, both analytical approaches, ISD fragmentation of protein, and analysis of PSD fragments of intense ladder ions in the 1 to 8 kD range, can be combined to extend a sequence tag toward the terminus and/or to increase confidence in O-glycosylation site assignment.

CONCLUSIONS

In ISD-MALDI the electron-transfer from hydrogen radicals of the matrix to singly charged analyte ion species that are prevalent in MALDI results in a significant reduction of sensitivity and necessitates the application of much higher protein amounts compared to PSD-MALDI or ETD-ESI on the peptide level. Attempts to increase ion yields in the positive ion mode by terminal fixed-charge tagging revealed significant decreases in fragmentation efficiency of standard peptides and glycopeptides. Further studies are needed to find conditions for achievement of higher fragment ion intensities. Such technical improvements could make the top-down sequencing of glycoproteins by ISD-MALDI-MS an attractive and powerful tool for the site determination of complex O-glycans at the protein level.

ASSOCIATED CONTENT

Supporting Information. Tables S1 to S5B. Lists of annotated ion signals. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: +49 221 478 4493. Fax: +49 221 478 7788. E-mail: franz.hanisch@uni-koeln.de.

ACKNOWLEDGMENT

This investigation was supported by grants of the Deutsche Forschungsgemeinschaft Ha 2092/21-1 and SFB 635. The author acknowledges technical assistance by the Central Bioanalytics of the Center for Molecular Medicine Cologne (Dr. Stefan Müller, Ursula Cullmann, Andrea Grün) and thanks Dr. Tilo Schwientek for providing the fusion protein MUC1-VH.

REFERENCES

(1) Julenius, K.; Mølgaard, A.; Gupta, R.; Brunak, S. *Glycobiology* **2005**, *15*, 153–164.

(2) Goletz, S.; Thiede, B.; Hanisch, F.-G.; Schultz, M.; Peter-Katalinic, J.; Müller, S.; Seitz, O.; Karsten, U. *Glycobiology* **1997**, *7*, 881–896.
(3) Hanisch, F.-G.; Green, B. N.; Bateman, R.; Peter-Katalinic, J. *J. Mass Spectrom.* **1998**, *33*, 358–362.
(4) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2005**, *4*, 628–632.
(5) Darula, Z.; Medzihradszky, K. F. *Mol. Cell. Proteomics* **2009**, *8*, 2515–2526.
(6) Brown, R.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 3990–3999.
(7) Suckau, D.; Cornett, D. S. *Anal. Mag.* **1998**, *26*, 18–21.
(8) Suckau, D.; Resemann, A. *Anal. Chem.* **2003**, *75*, 5817–5824.
(9) Resemann, A.; Wunderlich, D.; Rothbauer, U.; Warscheid, B.; Leonhardt, H.; Fuchser, J.; Kuhlmann, K.; Suckau, D. *Anal. Chem.* **2010**, *82*, 3283–3292.
(10) Resemann, A., Suckau, D. Automated acquisition of MALDI-ISD spectra for the N- and C-terminal sequence determination of intact proteins, Bruker Daltonics, Technical Note #TN-36, 2009
(11) Schwientek, T.; Mandel, U.; Roth, U.; Müller, S.; Hanisch, F.-G. *Proteomics* **2007**, *7*, 3264–3277.
(12) Müller, S.; Hanisch, F.-G. *J. Biol. Chem.* **2002**, *277*, 26103–26112.
(13) Engelmann, K.; Kinlough, C. L.; Müller, S.; Razawi, H.; Baldus, S. E.; Hughey, R. P.; Hanisch, F.-G. *Glycobiology* **2005**, *15* (11), 1111–1115.
(14) Pisano, A.; Packer, N. H.; Redmond, J. W.; Williams, K. L.; Gooley, A. A. *Glycobiology* **1994**, *4*, 837–844.
(15) Pisano, A., Jardine, D. R., Packer, N. H., Farnsworth, V., Carson, W., Cartier, P., Redmond, J. W., Williams, K. L., Gooley, A. A. In *Techniques in glycobiology, Identifying sites of glycosylation in proteins*; Townsend, R. R., Hotchkiss, A. T., Jr., Eds.; Marcel Dekker: New York, 1996; pp 299−320.
(16) Huang, Z.-H.; Shen, T.; Wu, J.; Gage, D. A.; Watson, J. T. *Anal. Biochem.* **1999**, *268*, 305–317.
(17) Rademaker, G. J.; Pergantis, S. A.; Blok-Tip, L.; Langridge, J. I.; Kleen, A.; Thomas-Oates, J. E. *Anal. Biochem.* **1998**, *257*, 149–160.
(18) Hanisch, F.-G.; Jovanovic, M.; Peter-Katalinic, J. *Anal. Biochem.* **2001**, *290*, 47–59.
(19) Mirgorodskaya, E.; Hassan, H.; Clausen, H.; Roepstorff, P. *Anal. Chem.* **2001**, *73*, 1263–1269.
(20) Zheng, Y.; Guo, Z.; Cai, Z. *Talanta* **2009**, *78*, 358–363.
(21) Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol. Cell Proteomics* **2002**, *1*, 791–804.
(22) Mirgorodskaya, E.; Hassan, H.; Wandall, H. H.; Clausen, H.; Roepstorff, P. *Anal. Biochem.* **1999**, *269*, 54–65.