Journal of Proteomics 134 (2016) 93–101

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Comprehensive glycosylation profiling of IgG and IgG-fusion proteins by top-down MS with multiple fragmentation techniques

Bao Quoc Tran [a], Christopher Barton [b], Jinhua Feng [b], Aimee Sandjong [b], Sung Hwan Yoon [a], Shivangi Awasthi [a], Tao Liang [a], Mohd M. Khan [a], David P.A. Kilgour [a], David R. Goodlett [a,*], Young Ah Goo [a,*]

[a] Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA
[b] MedImmune, LLC, Gaithersburg, MD, USA

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form 18 September 2015
Accepted 15 October 2015
Available online 17 October 2015

Keywords:
Top-down
Middle-down
Glycosylation
Protein species
MALDI-ISD
Oligosaccharide profiling

ABSTRACT

We employed top- and middle-down analyses with multiple fragmentation techniques including electron transfer dissociation (ETD), electron capture dissociation (ECD), and matrix-assisted laser desorption ionization in-source decay (MALDI-ISD) for characterization of a reference monoclonal antibody (mAb) IgG1 and a fusion IgG protein. Fourier transform ion cyclotron resonance (FT-ICR) or high performance liquid chromatography electrospray ionization (HPLC-ESI) on an Orbitrap was employed. These experiments provided a comprehensive view on the protein species; especially for different glycosylation level in these two proteins, which showed good agreement with oligosaccharide profiling. Top- and middle-down MS provided additional information regarding glycosylation sites and different combinational protein species that were not available from oligosaccharide mapping or conventional bottom-up analysis. Finally, incorporating a limited enzymatic digestion by immunoglobulin G-degrading enzyme of *Streptococcus pyogene* (IdeS) with MALDI-ISD analysis enabled extended sequence coverage of the internal region of protein without pre-fractionation.

Biological significance: Oligosaccharide profiling together with top- and middle-down methods enabled: 1) detection of heterogeneous glycosylated protein species and sites in intact IgG1 and fusion proteins with high mass accuracy, 2) estimation of relative abundance levels of protein species in the sample, 3) confirmation of the protein termini structural information, and 4) improved sequence coverage by MALDI-ISD analysis for the internal regions of the proteins without sample pre-fractionation.

1. Introduction

Protein top-down analysis by mass spectrometry (MS) has provided efficient characterization of proteins in their intact forms including any "protein species" [1–4]. According to Schlüter and Jungblut et al. [3], protein species are defined as the different proteins derived from a single gene. Post-translational modifications (PTM), nucleotide polymorphism, and alternative splicing can generate a large number of different protein species that derive from a single gene. Unlike protein isoforms, which are genetically defined, a protein species is chemically defined and thus includes also recombinant proteins such as monoclonal antibodies with different PTMs. Top- and middle-down approaches could provide complementary information to the bottom-up analysis of protein species because bottom-up analysis alone, in which peptides from all of the different protein species of a protein are mixed, can lead to inconclusive detection of some of the protein species present in a mixture. Analysis of intact proteins provides additional capabilities to identify PTMs such as glycosylation that are commonly found in mAbs. Recently, incorporation of tandem mass spectrometry (MS/MS) techniques like electron transfer dissociation (ETD) and higher collisional energy dissociation (HCD) has improved sequence coverage by top-down analysis [5,6]. Proteins can also be studied further by limited or specific proteolysis to yield information about protein isoforms and PTMs via analysis of large peptides or protein fragments, which has been referred to as middle-down analysis.

Antibodies represent a class of proteins that are key to mammalian immunological defense systems. They bind to an antigen (protein, glycoprotein, DNA, etc.) with a high degree of specificity thus neutralizing foreign biological material and xenobiotics. This process makes them extremely valuable for use in diagnostics, general research, and as therapeutics. Since their introduction in the late 1980s, therapeutic mAbs have become very popular drug candidates due to this high specificity and low toxicity. At present, approximately 30 mAbs are used for human therapeutics for oncology, autoimmunity, inflammation, infectious diseases, and metabolic disorders [7,8]. Most of the clinically approved mAbs are IgG1s, which are composed of two heavy chains (50 kDa each) and two light chains (25 kDa each) covalently linked by disulfide bridges. Carboxyl-terminal domains (Fc domains) are conserved, but amino-terminal domains (Fv domains) are variable in amino acid sequence,

* Corresponding authors.
E-mail addresses: dgoodlett@rx.umaryland.edu (D.R. Goodlett), ygoo@rx.umaryland.edu (Y.A. Goo).

http://dx.doi.org/10.1016/j.jprot.2015.10.021
1874-3919/

resulting in variations in specificity. The structure of IgGs, with four high molecular weight polypeptide chains held by inter- and intra-disulfide bonds as well as non-covalent interactions make characterization of mAbs by top down mass spectrometry very challenging.

As a complementary technique, middle-down analysis of the fragments of mAbs produced by specific enzymes such as papain, pepsin or immunoglobulin G-degrading enzyme of *Streptococcus pyogenes* (IdeS) [9] can facilitate characterization. For instance, analyzing Fab and $F(ab')_2$ fragments offer advantages over intact mAb analysis by reducing nonspecific antigen binding to Fc region. On the other hand, the Fc region can be used for in-depth glycosylation analysis.

As contribution to the special issue "Protein Species", we sought to study the sequence and PTM status of IgG molecules by top- and middle-down MS. A protein species is defined by chemical modifications (i.e. sequence chain length and/or PTM) and includes recombinant proteins such as mAbs produced as drugs [1–3]. As such, we employed oligosaccharide profiling analysis for comprehensive glycan identification of an IgG1 mAb and a glycosylated IgG1 fusion protein. This was used as an orthogonal approach to top- and middle-down analyses with multiple fragmentation techniques including electron capture dissociation (ECD) [10], electron transfer dissociation (ETD) [11], and MALDI-in-source decay (ISD) [12]. All these fragmentation methods were employed on one of two high-resolution MS instruments, a Fourier transform ion cyclotron resonance (FT-ICR) MS [13,14] or an Orbitrap Elite MS. FT-ICR data were further processed by application of a novel phase correction software, Autophaser [15]. We determined heterogeneous combinations of glycosylation on intact form of the proteins with high relevance to the glycan profiling. Pyroglutamate modification and lysine truncation were also shown to occur respectively on N-terminus and C-terminus of the proteins. In particular, we demonstrated the advantage of MALDI-ISD over ESI (electrospray ionization) for analysis of a protein with high sialic acid modification, and for improved sequence coverage by a single MALDI-ISD analysis of IdeS digest without sample pre-fractionation.

2. Material and methods

2.1. Protein samples

The recombinant human mAb (IgG1) and IgG1 fusion protein used in this study were manufactured using standard cell culture, purification and formulation processes at MedImmune LLC. The mAb is a novel IgG1 composed of two identical heavy and two identical light chains, linked by disulfide bonds, with an approximate molecular weight of 150 kDa. The IgG1 fusion protein is comprised of two identical chains with a human protein domain fused onto the constant domain (Fc) of a human IgG1 heavy chain. These two chains are covalently linked by disulfide bonds at the IgG1 hinge, with an approximate molecular weight of 90 kDa. The mAb possesses a single N-glycosylation site on each heavy chain within the CH2 domain (N297), while the fusion protein has three N-glycosylation sites; one on the CH2 domain of the IgG (N208) and two within the fusion protein (N76 and N108).

The mAb was expressed in stably transfected NS0 cells and purified using a standard multi-step purification process [16]. It is a fully human monoclonal antibody (IgG1). The product quality of the mAb was confirmed by non-reducing gel electrophoresis (96.0% purity) and high performance size exclusion chromatography (99.7% purity). The fusion protein was expressed in stably transfected Chinese hamster ovary cells (CHO) and purified using a standard multi-step purification process [16]. The product quality of the IgG fusion protein was confirmed by non-reducing gel electrophoresis (99.0% purity) and high performance size exclusion chromatography (99.8% purity).

2.2. Oligosaccharide profiling

Protein glycosylation was analyzed using a protocol modified from Ahn et al. [17]. Briefly, the protein was denatured in 2% *w*/*v* SDS, 1 M ß-Mercaptoethanol (product #63,689, Fluka, St. Louis, MO, USA) to ensure complete deglycosylation. The denatured fusion protein and the native mAb were then incubated with PNGase F (New England Bio Labs, Ipswich, MA) at 37 °C in 50 mM of Tris buffer pH 7.8 to remove all N-linked oligosaccharides. The released oligosaccharides were subsequently labeled at 37 °C for 60 min with a solution of 0.2 mM of 2-aminobenzamine (2-AB), 0.2 mM of picoline borane (PICB), and 7.5% formic acid (FA) in acetonitrile (ACN). The labeled glycans were reconstituted in 100% ACN and then purified by solid phase extraction to remove any large interfering molecules. The oligosaccharides were diluted into 100% ACN, separated by Acquity UPLC using BEH Amide column (130 Å, 1.7 μm, 2.1 mm × 150 mm). The UPLC chromatography was performed using 50 mM of ammonium formate (buffer A) and 100% ACN (buffer B) flowing at 0.5 mL/min (60 °C). For the mAb, a linear gradient over 23 min from 27% A to 41% A was used to elute the oligosaccharides, whilst a linear gradient from 28% A to 38% A over 45 min was used for the fusion protein. Glycan peaks were detected by fluorescence detection (Ex = 330, Em = 420) and identified by retention time with reference to previously characterized standards. Each peak was quantified based on the individual peak area relative to the total peak area detected. Peaks below the quantification limit of the assay (2%) were reported as 'minor' and those below the detection limit (0.1%) were reported as 'not detected' (ND).

2.3. Sample preparation for mass spectrometry (MS) analysis

At first, the samples were treated to remove impurities that were present in the sample storage buffer prior to the mass spectrometry analysis by off-line buffer exchange to 5% ACN and 0.1% TFA using Amicon 100 kDa or 50 kDa molecular weight cutoff (MWCO) centrifugal filters (product # UFC510024 and UFC505024, Merck Millipore, Tullagreen, Carrigtwohill, Ireland) for the mAb and the fusion protein, respectively. For the fusion protein that was in detergent-containing buffer, additional detergent removal was applied to the sample by Pierce detergent removal kit (product # 87778, Thermo Scientific, Rockford, CA, USA) following the manufacturer's instruction. Purified intact proteins were diluted in 50% acetonitrile and 0.1% formic acid at a final concentration of 12.5 μg/μL (mAb) and 10 μg/μL (fusion protein) for the top-down MS analysis.

To reduce protein by breaking off disulfide bonds, the samples were diluted to 1 μg/μL in 50 mM of ammonium bicarbonate buffer prior to dithiothreitol (DTT, product #43815, Sigma-Aldrich, St. Louis, MO, USA) addition at final concentration of 20 mM and incubated at 80 °C for 15 min. The reduced proteins were concentrated using Amicon 10 kDa centrifugal filter (product # UFC201024, Merck Millipore). The fusion protein was then subjected to the detergent removal using Pierce detergent removal kit (Thermo scientific). The reduced proteins were diluted in 50% ACN and 0.1% FA at a final concentration of 17 μg/μL (mAb) and 7 μg/μL (fusion protein) for the top-down MS analysis. The reduced proteins were also used for the MALDI-ISD application. The matrix 1,5-diaminonaphthalene (1,5-DAN, product # 56451, Fluka, St. Louis, MO, USA) in ACN was prepared fresh. One microliter reduced protein was spotted first and 1 μL matrix was added to the protein spot on a stainless steel polished MALDI plate and then dried.

For the middle-down, proteins were digested by an enzyme, IdeS (product # A0-FR1-020, Genovis AB, Lund, Sweden), according to manufacturer's instruction. Briefly, the proteins were diluted to 1 μg/μL in 50 mM of ammonium bicarbonate solution. The IdeS was added at a ratio of 1 unit per 1 μg of protein followed by 2 h of incubation at 37 °C. For the IgG1 mAb, additional filtration was applied to separate Fc and Fab fractions using a 50 kDa spin filter (Merck Millipore). To further reduce the proteins, DTT was added to final concentration of 20 mM to cleave disulfide bonds at 80 °C for 15 min. The reduced samples were concentrated using 10 kDa filter (Merck Millipore). The concentrated fusion protein was subjected to additional detergent removal using Pierce detergent removal spin column (Thermo Scientific).

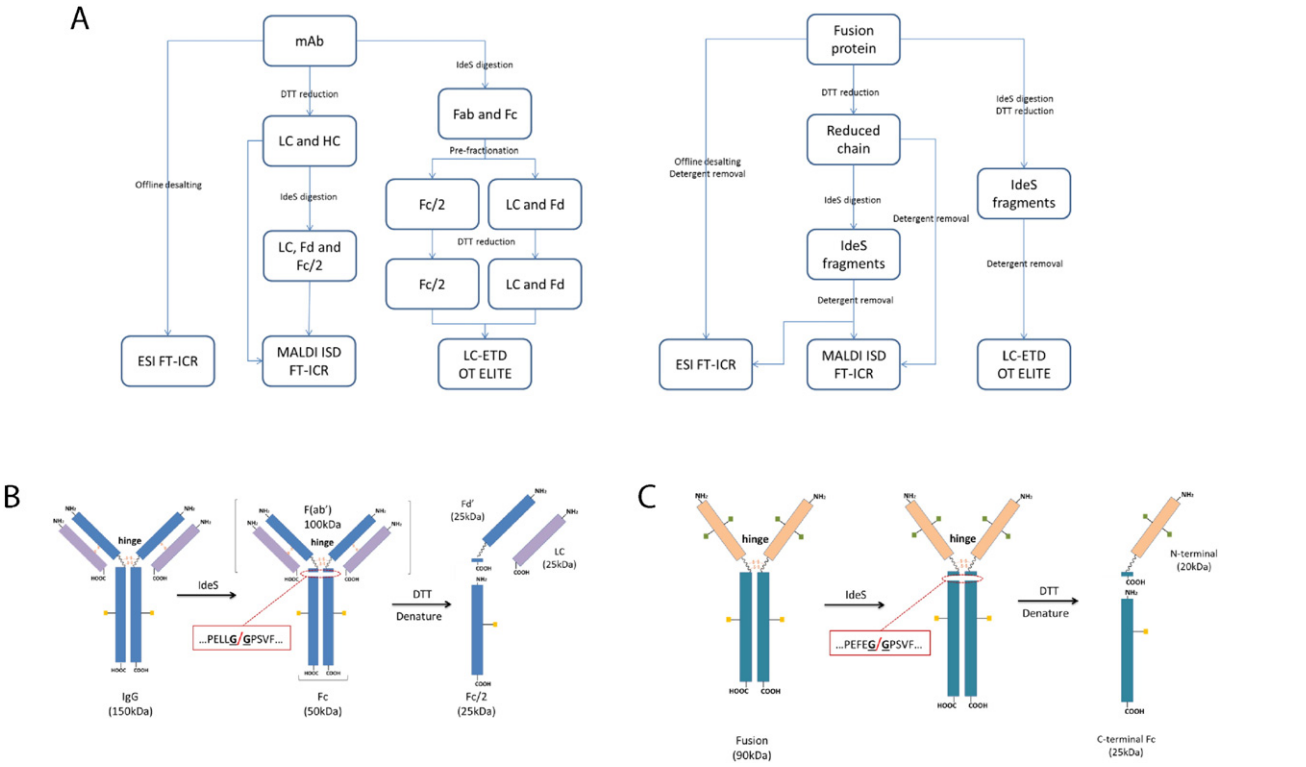

Fig. 1. Sample preparation and data acquisition workflow for (A) top-down and middle-down analyses of the IgG1 mAb and the fusion protein, (B) IdeS digestion of the mAb produces Fc and F(ab') domains which are further cleaved to Fc/2, LC and Fd' by DTT treatment, and (C) IdeS digestion of the fusion protein cleaves right below the hinge region and further releases N-terminal and C-terminal Fc domains by DTT treatment.

Prior to MS analysis, the samples were buffer exchanged to 5% ACN and 0.1% TFA. Approximately 7 μg/μL of the final sample concentration was used for the middle-down MS analysis.

2.4. MS data acquisition and data processing

Accurate mass measurement of the intact protein was acquired as follows. The purified intact protein was ionized by electrospray and introduced to 12 Tesla FT-ICR mass spectrometer (Solarix XR, Bruker, Bremen, Germany) at a flow rate of 2 μL/min. Accurate mass spectrum was recorded in magnitude mode in m/z range of 400–4000 at transient data of 128 k words with an accumulation of 1700 scans for the fusion protein and 2000 scans for the mAb. Continuous accumulation of selected ions (CASI) [18], a gas-phase ion enrichment technique, was applied to enhance the peak cluster signal at different charge states. CASI isolates ions of a narrow m/z window and stores them in an ion trap for repetitive cycles of laser ablation prior to sending the enriched ion population to the ICR cell for detection. The technique increases the dynamic range for peak detection. For an example, CASI in range of 2340–2380 m/z for charge state 63+ of the intact mAb was acquired. The reduced and IdeS treated fusion protein was also measured in CASI for charge state 17+. FT-ICR data was processed using DataAnalysis software version 4.2 (Bruker). Mass and charge state of the protein species were manually calculated with Charge State Ruler function.

ETD of the reduced proteins was carried out by isolating a single charge state in collision cell while ECD was performed for several selected charge states in ICR cell. ETD reagent accumulation time, ion accumulation time and reaction time were tuned for optimal fragment ion signal.

For the MALDI-ISD FT-ICR, in-source decay (ISD) spectrum was acquired with a transient of 1 mega (M)-word data points, over a 400–10,000 m/z with estimated resolving power of 270,000 at m/z 400. Quadrupole filtration was set at 1000 or 2000. Magnitude mode spectra of MALDI-ISD experiments were subjected to Hanning algorithm based baseline and phase correction using Autophaser [15]. Sophisticated numerical annotation procedure (SNAP) algorithm [19] was applied to the phased MALDI-ISD spectrum as well as ETD and ECD spectra extracting isotopic fragment ions at a quality factor threshold of 0.5 and signal-to-noise of 0.5, and relative intensity threshold (base peak) of 0.01%. The phased spectrum was post-acquisition calibrated using the observed c-ions as internal calibrants. c-, z- and z + 2 ions were identified by matching the observed masses with theoretical ones using Biotools version 3.2 (Bruker) and Sequence Editor 3.2 (Bruker) at a MS/MS mass tolerance of 50 ppm. G0f and G1f glycosylation of serine site 300, protein N-terminal modifications were defined as optional modifications.

The IdeS treated and reduced mAb was also subject for the LC ETD MS/MS acquisition on Orbitrap Elite (Thermo Fisher Scientific, Saint Jose, CA, USA). Approximately14 μg protein sample containing the Fc

Table 1
Glycosylated protein species and relative abundance level detected by oligosaccharide profiling of the IgG1 and the fusion protein. Glycans in bold are also detected by top-down and middle-down analysis. Peaks below the quantification limit of the fluorescence detection assay (2%) were reported as 'Minor' and peaks below 0.1% detection limit were reported as 'not detected' (ND). Additional information regarding the glycan nomenclature and structure is provided in Ref. [21], Fig. 5.

Glycan	mAb	Fusion protein	Glycan	mAb	Fusion protein
G0-GN	Minor	ND	G2f + G	>**2%**	ND
G0f-GN	>**2%**	**Minor**	G1f + NGNA	Minor	ND
G0	Minor	>**2%**	G2f + 2G	**Minor**	ND
G0f	>**30%**	>**10%**	G2f + NGNA	Minor	ND
M5	Minor	>**2%**	G2f + G + NGNA	Minor	ND
G1f-GN	Minor	ND	G1f + NANA	ND	>**2%**
G1	ND	Minor	G2f + NANA	ND	>**10%**
G1f	>**25%**	>**2%**	G2f + 2NANA	ND	>**10%**
M6	**Minor**	ND	G3f + NANA	ND	>**2%**
G1f-GN + G	Minor	ND	G3f + 2NANA	ND	>**2%**
G2f	>**5%**	>**2%**	G3f + 3NANA	ND	>**2%**
G1f-GN + NGNA	Minor	ND			

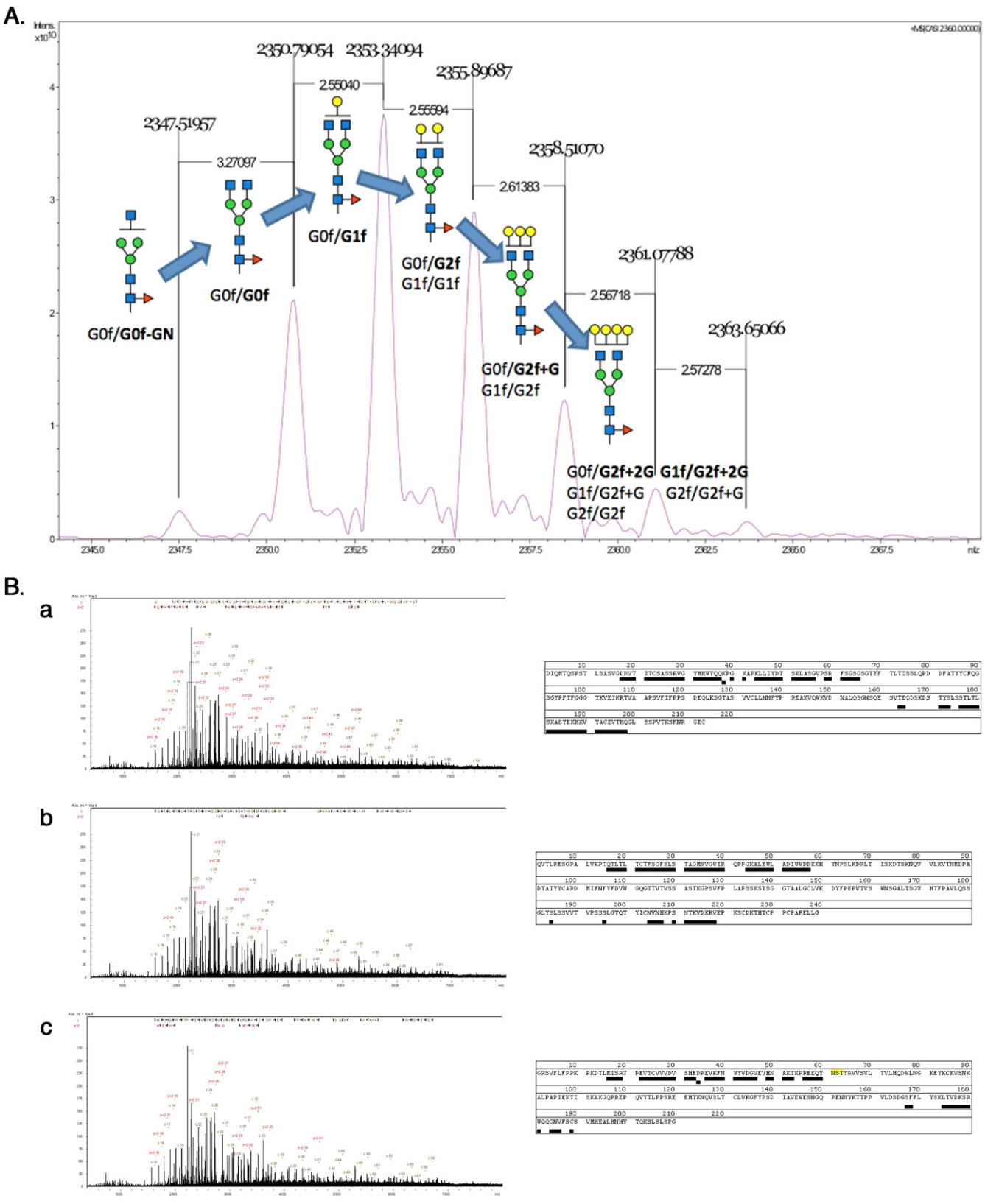

Fig. 2. A. Mass spectrum of the 63 + charge state of IgG1 mAb generated by continuous accumulation of selected ion (CASI) mass spectrometry at m/z 2360. The mass spectrum was result from accumulation of 2000 scans collected in a time domain of 128 K word data points. The peaks demonstrate heterogeneous glycosylation of protein species in combination with C-terminal lysine truncation and N-terminal glutamine (Q) to pyroglutamate (pyroGlu) conversion on the heavy chain. Bold font indicates the glycoforms identified by both intact protein analysis and oligosaccharide profiling analysis. B. MALDI-ISD mass spectrum of the DTT/IdeS treated IgG1 mAb. (a) light chain, (b) Fd, and (c) Fc/2 regions. Amino acid residues underlined were detected by c or z+2 ions after phase correction with mass error threshold of 50 ppm, covering 33.8% of the light chain, 24.8% Fd, and 22.6% Fc/2. N-terminal pyroglutamate (pyroGlu, pE) residue and C-terminal lysine truncation were detected on the heavy chain.

Table 2
Glycosylated protein species detected by accurate mass measurement of the IgG1 mAb by top- and middle-down analysis of 63+ charge state. Bold font indicates the glycoforms identified by both intact protein analysis and oligosaccharide profiling analysis. The abundance level of the glycoforms was in good agreement by the two methods. Additional information regarding the glycan nomenclature and structure is provided in Ref. [21], Fig. 5.

Glycoforms	Peak	z	Observed mass	Theoretical mass	Mass error (Da)	Mass error (ppm)
G0f/G0f-GN	2347.51957	63+	147,830.27476	147,831.977282	−1.70	−11.5
G0f/G0f	2350.79054	63+	148,036.34576	148,035.056655	1.29	8.7
G0f/G1f; G0f/M6	2353.34094	63+	148,197.0266	148,197.109478	−0.08	−0.5
G0f/G2f; G1f/G1f	2355.89687	63+	148,358.04457	148,359.162301	−1.11	−7.5
G0f/G2f + G; G1f/G2f	2358.51070	63+	148,522.71564	148,521.215124	1.50	10.1
G0f/G2f + 2G; G1f/G2f +G; G2f/G2f	2361.07788	63+	148,684.44818	148,683.267947	1.18	7.9
G1f/G2f + 2G; G2f/G2f + G	2363.65066	63+	148,846.53319	148,845.32077	1.21	8.1

or Fab fractions was loaded on a 100 μm × 20 mm column packed with Jupiter 5 μm C5 particles (product # 04A-4052, Phenomenex, Torrance, CA, USA). The protein fragments were eluted by a fast gradient of ACN from 5% to 50% in 15 min at a flow rate of 0.3 μL/min. The fragments were scanned in range of 350–3500 m/z at resolving power of 120,000. Most intense peaks were separated and selected for ETD fragmentation. ETD reaction time was optimized for an individual precursor peak. ETD spectrum was acquired with 5 microscans at resolving power of 120,000 and the averaged scan was used for the sequencing by MASH Suite version 1.0 [20].

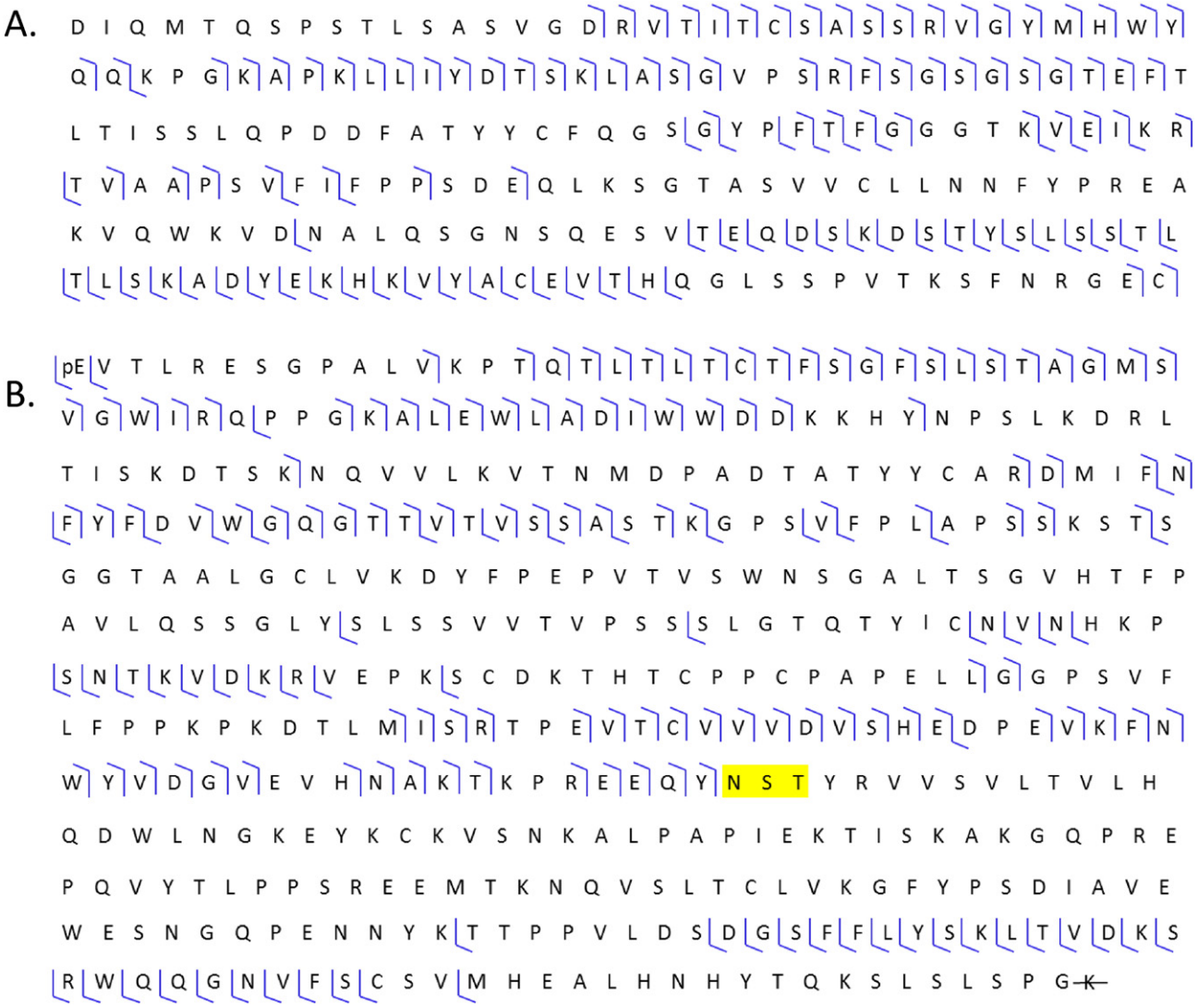

Fig. 3. Amino acid sequence coverage for (A) IgG1 mAb light chain and (B) heavy chain determined by a combination of MALDI-ISD and middle-down ETD analyses. The overall sequence coverage was 35.9% and 52.6% for the heavy and light chains, respectively. Glutamine (Q) to pyroglutamate (pyroGlu, pE) conversion and C-terminal lysine truncation were detected on the heavy chain. One glycosylation site is highlighted in yellow.

3. Results and discussions

3.1. Analysis of the IgG1 mAb

The various sample preparation workflows for MS based analyses of the IgG1 are provided in Fig. 1A. The IgG1 was analyzed in various forms including intact, reduced, and as IdeS produced fragments. These were subjected to high mass accuracy measurements of protein species. Oligosaccharide profiling by enzymatic release of glycans and HPLC separation provided a comprehensive glycan list (Table 1, see Figure 5 in ref. [21]). This approach, however, required multiple sample preparation steps that include enzymatic glycan cleavage by PNGase F and oligosaccharide labeling. Additionally, it could not resolve a combination of glycans present in different protein species, such as the glycosylation status of all polypeptide chains on an intact mAb. On the other hand, intact protein top-down analysis by ESI FT-ICR CASI at charge state 63+ successfully identified multiple glycoforms on the IgG1 (Fig. 2A). The accurate mass measurement enabled not only the determination of glycan

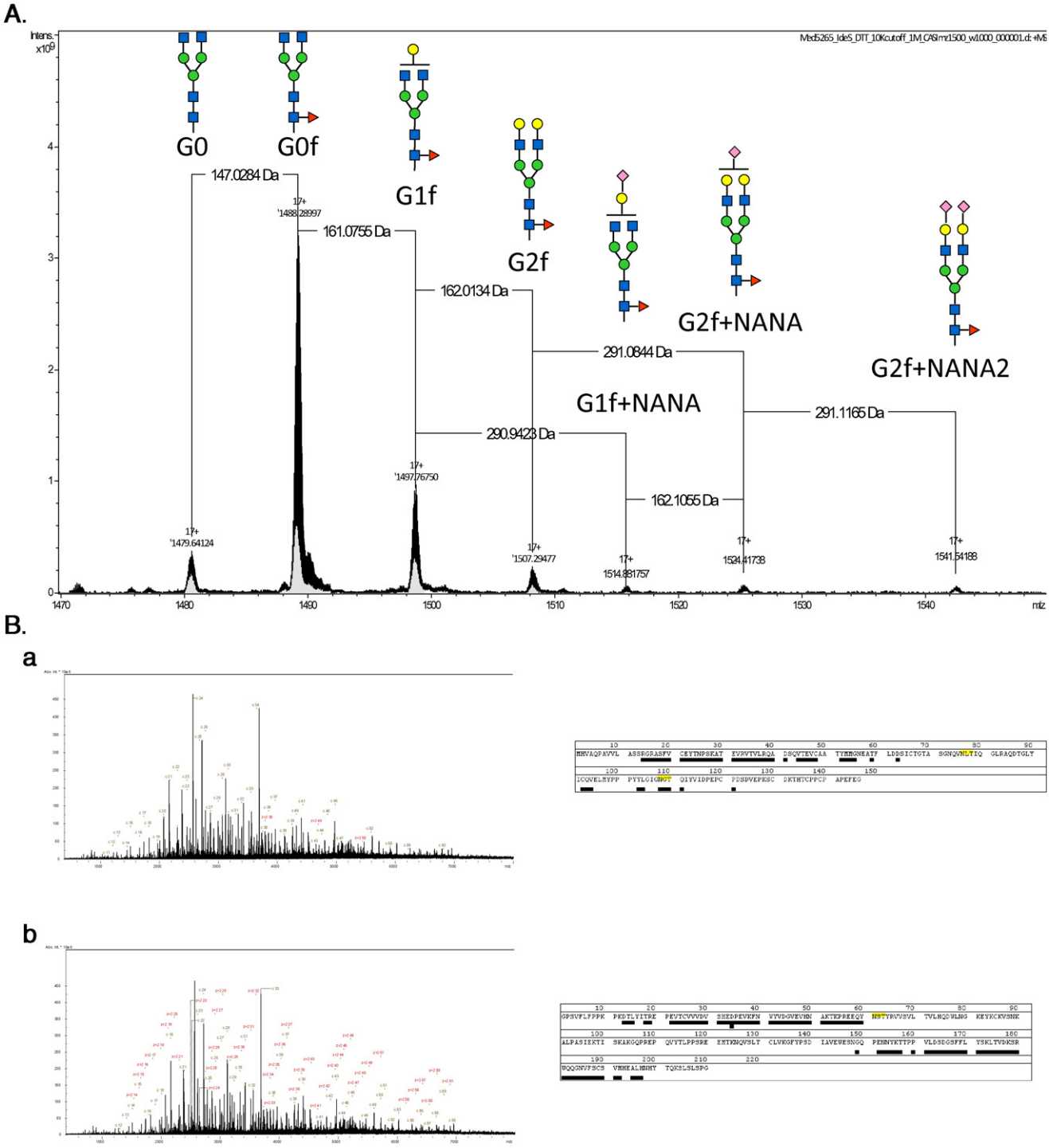

Fig. 4. A. Mass spectrum of the 17+ charge state of fusion protein generated by continuous accumulation of selected ion (CASI) mass spectrometry at m/z 1500. The mass spectrum was result from accumulation of 200 scans collected in a time domain of 1 M word data points with an estimated resolving power of 270,000 at m/z 400. Major peaks demonstrate heterogeneous combination of glycosylation with low levels of sialylation. B. MALDI-ISD mass spectra of the DTT/IdeS-treated fusion protein. The phase corrected mass spectrum was matched to sequences of (a) N-terminal and (b) C-terminal fragment. Annotated fragment ions, c and z+2, span 33.6% and 41.4% of N-terminal and C-terminal sequences, respectively.

combinations but also the relative abundance of the glycoforms as shown by the seven abundant peaks. The detected glycoforms demonstrated heterogeneous heavy chain glycosylation as detailed in Table 2. The seven detected glycoforms represented the most abundant seven out of the 17 glycans determined by oligosaccharide profiling (Table 1). And the abundance levels of the identified glycoforms were in good qualitative agreement between the two analyses, in which the combination of G0f and G1f represented the most abundant glycoform observed.

Modifications on termini, such as C-terminal lysine truncation and/or N-terminal glutamine (Q) to pyroglutamate (pyroGlu) conversion, are other PTMs that could occur on mAbs during production or storage process [22,23]. The seven most abundant protein species of the IgG1 were determined to contain N-terminal pyroGlu conversion and C-terminal lysine truncation on the heavy chain at average mass error 0.3 Da (2.1 ppm).

Analyses of the reduced IgG1 using the ESI FT-ICR and LC–MS on OT-Elite was performed to determine if the different glycoforms can be detected and the termini modification can be confirmed. The observed accurate masses of the heavy chains indicated predominant glycosylation of G0f, G1f and G2f as well as the truncation of C-terminal lysine and N-terminal pyroGlu conversion (data not shown). The results were in good agreement with the intact analysis of the IgG1.

MALDI-ISD has been developed and used as an alternate to N-terminal Edman sequencing for protein characterization [12]. The advantage of MALDI-ISD is the rapid sequencing capability of both N- and C-terminal for greater than 50 amino acid residues without the need for protein hydrolysis. In comparison, Edman sequencing is limited to N-terminal sequences of polypeptides shorter than 50 residues (in practice, under 30) due to the efficiency of release of the amino acid derivative [24]. We employed 1,5-DAN as MALDI matrix in our experiments since it was reported to be one of the most efficient matrices for MALDI-ISD to generate c- and z + 2 ion series [25,26]. To improve fragmentation efficiency for better sequence coverage, the IgG1 was incubated with DTT to break disulfide bonds producing heavy and light chains prior to MALDI-ISD analysis, but these were not separated prior to MS analysis. The phase corrected MALDI-ISD mass spectrum of the reduced IgG1 (see Figure 1 in Ref. [21]) was searched against a light chain or heavy chain sequence database with the optional C-terminal lysine truncation and N-terminal pyroGlu conversion. The notable observation was the detection of up to 70 N-terminal residues and 47 C-terminal residues for the light chain whereas the heavy chain sequence was detected by slightly fewer residues. Phase correction of the MALDI-ISD spectrum by Autophaser software improved the resolving power and mass accuracy of the peaks, but showed no significant improvement in detected sequence coverage.

Given the typical 400–10,000 m/z range that MALDI-ISD covers, obtaining a complete sequence of the heavy (349 aa) and light chains (213 aa) was quite challenging especially for the internal sequence regions. This challenge prompted the use of the immunoglobulin degrading enzyme from *Streptococcus pyogenes*, IdeS [9], which cleaves Glycine–Glycine (G–G) located underneath the IgG1 hinge region to further fragment the heavy chain into the Fc/2 and Fd prior to MALDI-ISD analysis (Fig. 1B). The MALDI-ISD of IdeS treated sample produced more product ions from simultaneous fragmentation of the Fc/2 and Fd regions and resulted in improved sequence coverage for the heavy chain (Fig. 2B). The overall sequence coverage by MALDI-ISD was 37.6% and 23.6% for the light and heavy chains, respectively. MALDI-ISD has been previously applied to top-down PTM studies for O-glycosylation and other glycoproteins [26,27]. Similar to the observation made by Chaurand et al. [28], we noted that the detection of the c-ions of mAb Fd domain in the MALDI-ISD mass spectrum stopped at Y299, which is right before NST motif of N-glycosylation (Fig. 2B). This observation suggests that the peptide bond fragmentation could be hampered in the presence of the carbohydrate group.

In recent reports, IdeS treated protein fragments were LC separated to produce light chain, Fd and Fc/2 prior to either ETD analysis or MALDI-ISD using super DHB matrix [9,29]. LC-ETD analysis of Adalimumab resulted in sequence coverage of 50% for Fc/2 and light chain, and 30% for Fd domains [9]. LC-MALDI-ISD of off-line pre-fractionated centuximab produced 61.5% light chain coverage while no information regarding heavy chain coverage was provided [29]. In the current study, MALDI-ISD analysis of IdeS treated mAb fragments of approximately 25 kDa allowed sequence up to 37.6% for the light chain, 22.6% for Fd and 24.8% for Fc/2 in a single analysis without need for pre-fractionation or combing a multiple off-line LC fractions.

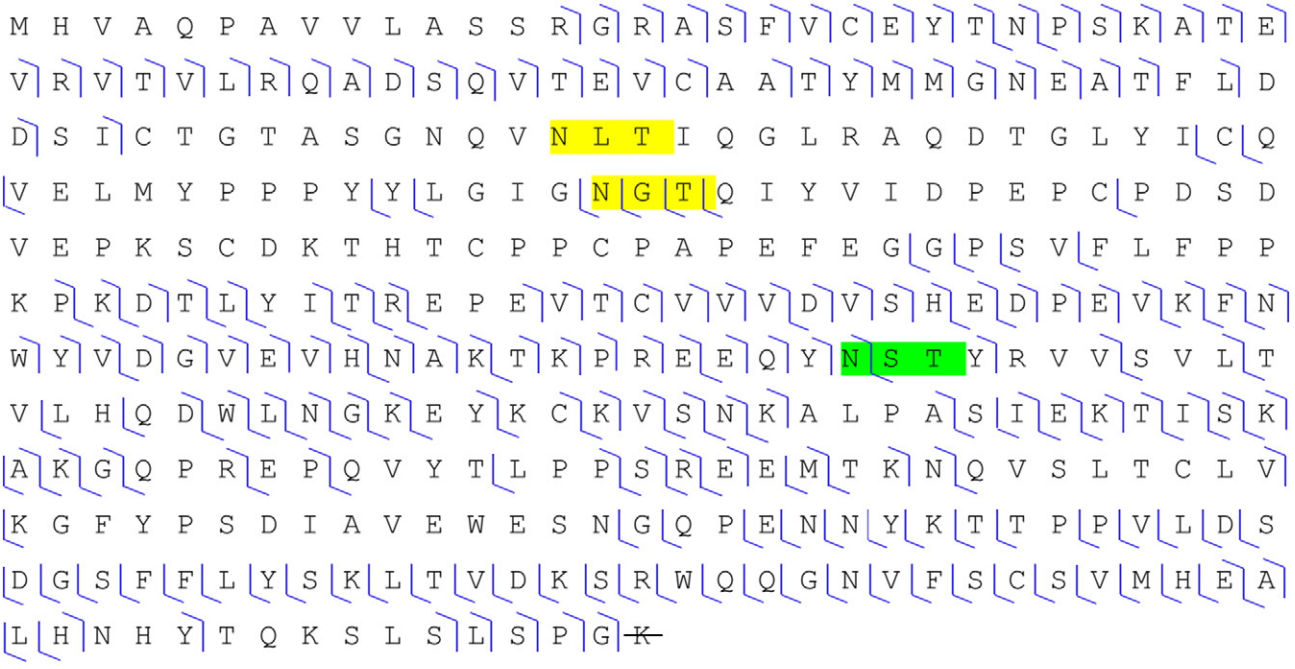

Fig. 5. Fusion protein sequence coverage was determined by combination of MALDI-ISD and middle-down ETD analysis with an overall sequence coverage of 61.0%. C-terminal lysine truncation and three glycosylation sites are highlighted (green-located on Fc; yellow-located on non-Fc).

The IdeS treated mAb sample was further subject to on-line clean up by rapid nano-LC gradient coupled to an OT-Elite. The precursor ions of each fraction, including the light chain, Fc/2, and Fd, were isolated and subject to ETD fragmentation generating c- and z- ions. Analysis of the ETD tandem mass spectra resulted in detection of fragment ions that cover the central regions, 92–122 for the light chain and 98–138 for the Fd fragment (see Figure 3A and 3B in ref. [21]). These central regions were previously not accessible by MALDI-ISD analysis of the reduced sample. Additionally, N-terminal glutamine to pyroGlu conversion of the heavy chain as previously determined by the intact protein analysis was confirmed with 3.9 ppm error (0.1 Da) from ETD analysis of the Fd domain (see Figure 3B in Ref. [21]).

While the top-down analysis of the intact or reduced protein provided accurate and rapid determination of different glycosylation protein species of the mAb, the middle-down analysis by IdeS proteolysis greatly enhanced the sequence coverage by providing more efficient fragmentation capability. Seven highly abundant glycosylated protein species were determined which were all modified with N-terminal pyroGlu conversion and C-terminal lysine truncation of the heavy chain. The MALDI-ISD showed a superior performance for sequencing over ECD and ETD analysis in the FT-ICR (data not shown). The middle-down analysis following IdeS by LC–MS/MS ETD greatly enhanced sequencing of the central regions of the mAb. By employing multiple fragmentation techniques, a sequence coverage of 35.9% and 52.6% for heavy and light chains, respectively, was obtained (Fig. 3). Difficulty in accessing the internal region of the sequence might relate to reformation of disulfide bonds that would prevent effective fragmentation of the protein, especially from ESI generated ions. However, even with alkylation of cysteine and use of other denaturing chemicals, such as guanidine hydrochloride we did not see improvement in sequence coverage (data not shown). Both the size and structural conformation of the studied IgG1 could be a contributing factor as well. Considering a recent study that reported 100% sequence coverage of mAb by analyzing peptide fragments as small as 3.4 kDa on average by extended bottom-up approach [30], the sequence coverage observed in this study from analyzing 25 kDa fragment was reasonable.

3.2. Analysis of the IgG fusion protein

The pattern of glycosylation in an IgG1 with a single N-glycosylation site in the Fc is less complex than that observed on proteins with non-Fc glycosylation sites. In order to test the analytical approach on a protein with non-Fc glycosylation, we analyzed the sequence of an IgG fusion protein. The fusion protein is a dimer comprised of six total N-glycosylation sites (one Fc and two non-Fc sites within the fusion protein per chain) with a heterogeneous range of glycan structures as shown by the oligosaccharide mapping analysis (Table 1). These include bi- and tri-antennary sialylated species not normally seen on Fc glycans and likely arising from the two glycosylation sites within the fusion protein. Intact protein analysis by FT-ICR supported the oligosaccharide data, detecting a complex combination of glycosylated protein species by use of only accurate mass measurement (see Figure 4 in Ref. [21]). These glycoforms were predominantly sialic acid modifications determined by mass difference of 37 on protein species at charge state 38+ of the fusion protein. The most abundant protein species observed had the combination of $(G1f)_1$, $(G0f)_1$, $(G2fS1)_2$, and $(G2fS2)_2$ with C-terminal lysine truncation, which was determined with a mass error of 1.6 Da (18 ppm). Due to the large number of heterogeneous glycoforms in the intact sample, which hampered accurate detection of each glycoform, consecutive analysis was done by use of CASI FT-ICR after DTT reduction and/or on a DTT/IdeS treated sample. Use of IdeS cut the reduced fusion protein into two large peptides of approximate 45 kDa each, resulting in one N-terminal domain fragment and a C-terminal Fc fragment (Fig. 1C). By analyzing the C-terminal fragment, seven glycoforms in which G0f and G1f were the most high abundant were observed (Fig. 4A). The detailed mass measurement and error information of these glycoforms can be found in Ref. [21], Table 1. This data was in agreement for the number of glycans detected by oligosaccharide profiling (in Table 1) and demonstrates the ability of middle-down MS to determine the site, as well as the nature, of glycosylation on a protein.

The N-terminal fragment of the fusion protein has two N-glycosylation sites per chain, however ESI analysis by FT-ICR or OT-Elite could not detect the presence of the N-terminal fragment suggesting poor ionization due to the structural conformation, high level of negative charge sialic acid residues or stoichiometric properties of the N-terminal fragment. The large number of heterogeneous glycosylation might also contribute to the poor signal with low detection limit.

MALDI-ISD was employed to analyze the DTT reduced fusion protein with and without IdeS treatment. Similar to the IgG1 analysis, the phase corrected MALDI-ISD of the DTT reduced fusion protein confirmed the C-terminal lysine truncation and provided fragment ions that covered approximately 65 residues from both termini with a total of 25.3% sequence coverage leaving the most of the internal region not sequenced (see Figure 6 in ref. [21]). When IdeS digestion was added, the sequence coverage of the fusion protein increased to 38.2% (Fig. 4B). Notably, both the N- and C-terminal fragments of the fusion protein were detected by MALDI-ISD suggesting an advantage of this technique over ESI for analysis of fusion proteins with complex heterogeneous glycosylation sites.

Similar to the observation made on the mAb analysis, ISD fragmentation stopped right before the NST motif, at Y206, for the C-terminal fragment of the fusion protein. However, the NGT motif on the N-terminal fragment was observed by the z+2 ions suggesting that the NGT motif may have lost a glycan group during ISD fragmentation (Fig. 4B).

The DTT/IdeS treated sample was also subjected to LC-MS analysis on the Orbitrap Elite. Similar to the FT-ICR ESI analysis, only the C-terminal glycoforms were detected whereas none of the N-terminal peaks were present. Ionization efficiency of the N-terminal might be severely limited in both FT and OT instruments by the ESI. ETD mass spectra of the two most abundant peaks provided 47.6% sequence coverage for the C-terminal fragment and confirmed the G0f and G1f glycosylation via detection of glycan-containing fragment ions (see Figure 8 in Ref. [21]).

By use of the multiple fragmentation technique including MALDI-ISD and ETD, overall 61.0% sequence coverage was achieved (Fig. 5). The sequence analysis by MASH Suite indicated that the protein is truncated at K357, the C-terminus sequence. Comprehensive glycoforms were detected by intact protein analysis with less than 2 Da error in mass accuracy, which provided accurate identification of glycans, G0f, G1f and G0 (Fig. 4A). The abundance level of the each glycans was confirmed by analysis from ETD fragmentation. Finally, middle-down analysis of the DTT/IdeS treated sample provided detection of more than seven potential glycoforms with mass error less than 2.4 Da that was in good agreement with the oligosaccharide profiling data.

4. Conclusions

Top- and middle-down analyses, either used alone or in combination, together with multiple ion dissociation techniques proved in composite to be powerful tools for structural characterization of proteins circumventing what have otherwise been onerous sample preparation requirements. The top-down analysis by 12T FT-ICR and an Orbitrap provided superior performance in 1) detection of heterogeneous glycosylated protein species of intact proteins at high mass accuracy, 2) estimation of relative abundance levels of protein species, and 3) confirmation of the protein structural information (i.e. C-terminal K truncation and/or N-terminal modification). Middle-down analysis of the IdeS hydrolyzed protein fragments, particularly in combination with MALDI-ISD, enabled the confirmation of the N-terminal Glutamine to Pyroglutamate conversion for the mAb and C-terminal truncations in both mAb and the fusion proteins. It also enhanced sequencing of the

central region of the proteins resulting in improved sequence coverage. While MALDI-ISD showed limited performance determining N-glycosylation sites, use of MALDI in addition to ESI was critical in characterization of large proteins. This was due in part to the difficulty of keeping the two large protein fragments from reforming disulfide bonds that in turn affect structural conformation and ionization efficiency. Oligosaccharide mapping together with top- and middle-down methods enabled comprehensive glycan profiling and detection of the multiple glycosylation protein species in the proteins studied.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgments

This research was supported in part by collaborative grant from MedImmune, LLC (2014-2312 RSA), and University of Maryland, Baltimore, School of Pharmacy Mass Spectrometry Center (SOP1841-IQB2014). The authors would like to thank Dr. David Spencer for his critical reading of the manuscript.

References

[1] P. Jungblut, B. Thiede, U. Zimny-Arndt, E.C. Müller, C. Scheler, B. Wittmann-Liebold, A. Otto, Resolution power of two-dimensional electrophoresis and identification of proteins from gels, Electrophoresis 17 (1996) 839–847, http://dx.doi.org/10.1002/elps.1150170505.
[2] P.R. Jungblut, H.G. Holzhütter, R. Apweiler, H. Schlüter, The speciation of the proteome, Chem. Cent. J. 2 (2008) 16, http://dx.doi.org/10.1186/1752-153X-2-16.
[3] H. Schlüter, R. Apweiler, H.-G. Holzhütter, P.R. Jungblut, Finding one's way in proteomics: a protein species nomenclature, Chem. Cent. J. 3 (2009) 11, http://dx.doi.org/10.1186/1752-153X-3-11.
[4] L.M. Smith, N.L. Kelleher, Proteoform: a single term describing protein complexity, Nat. Methods 10 (2013) 186–187, http://dx.doi.org/10.1038/nmeth.2369.
[5] D.R. Ahlf, P.D. Compton, J.C. Tran, B.P. Early, P.M. Thomas, N.L. Kelleher, Evaluation of the compact high-field orbitrap for top-down proteomics of human cells, J. Proteome Res. 11 (2012) 4308–4314, http://dx.doi.org/10.1021/pr3004216.
[6] A. Chi, D.L. Bai, L.Y. Geer, J. Shabanowitz, D.F. Hunt, Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry, Int. J. Mass Spectrom. 259 (2007) 197–203, http://dx.doi.org/10.1016/j.ijms.2006.09.030.
[7] N.A.P.S. Buss, S.J. Henderson, M. McFarlane, J.M. Shenton, L. de Haan, Monoclonal antibody therapeutics: history and future, Curr. Opin. Pharmacol. 12 (2012) 615–622, http://dx.doi.org/10.1016/j.coph.2012.08.001.
[8] J.W. Larrick, Conference scene: progress with promising human antibodies, Immunotherapy 4 (2012) 257–261, http://dx.doi.org/10.2217/imt.12.5.
[9] L. Fornelli, D. Ayoub, K. Aizikov, A. Beck, Y.O. Tsybin, Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap fourier transform mass spectrometry, Anal. Chem. 86 (2014) 3005–3012, http://dx.doi.org/10.1021/ac4036857.
[10] R. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J. Am. Chem. Soc. 120 (1998) 3265–3266, http://dx.doi.org/10.1021/ja973478k.
[11] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9528–9533, http://dx.doi.org/10.1073/pnas.0402700101.
[12] D. Suckau, A. Resemann, MALDI top-down sequencing: calling N- and C-terminal protein sequences with high confidence and speed, J. Biomol. Tech. 20 (2009) 258–262.
[13] A.G. Marshall, C.L. Hendrickson, G.S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer, Mass Spectrom. Rev. 17 (1998) 1–35, http://dx.doi.org/10.1002/(SICI)1098-2787(1998)17:1<1::AID-MAS1>3.0.CO;2-K.
[14] E.N. Nikolaev, Y.I. Kostyukevich, G.N. Vladimirov, Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: theory and simulations, Mass Spectrom. Rev. (2014)http://dx.doi.org/10.1002/mas.21422.
[15] D.P.A. Kilgour, R. Wills, Y. Qi, P.B. O'Connor, Autophaser: an algorithm for automated generation of absorption mode spectra for FT-ICR MS, Anal. Chem. 85 (2013) 3903–3911, http://dx.doi.org/10.1021/ac303289c.
[16] P.A. Marichal-Gallardo, M.M. Álvarez, State-of-the-art in downstream processing of monoclonal antibodies: process trends in design and validation, Biotechnol. Prog. 28 (2012) 899–916, http://dx.doi.org/10.1002/btpr.1567.
[17] J. Ahn, J. Bones, Y.Q. Yu, P.M. Rudd, M. Gilar, Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 μm sorbent, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 878 (2010) 403–408, http://dx.doi.org/10.1016/j.jchromb.2009.12.013.
[18] M. Witt, J. Fuchser, B.P. Koch, Fragmentation studies of fulvic acids using collision induced dissociation fourier transform ion cyclotron resonance mass spectrometry, Anal. Chem. 81 (2009) 2688–2694.
[19] M.W. Senko, S.C. Beu, F.W. McLaffertycor, Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions, J. Am. Soc. Mass Spectrom. 6 (1995) 229–233, http://dx.doi.org/10.1016/1044-0305(95)00017-8.
[20] H. Guner, P.L. Close, W. Cai, H. Zhang, Y. Peng, Z.R. Gregorich, Y. Ge, MASH suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization, J. Am. Soc. Mass Spectrom. 25 (2014) 464–470, http://dx.doi.org/10.1007/s13361-013-0789-4.
[21] B.Q. Tran, C. Barton, J. Feng, A. Sandjong, S.H. Yoon, S. Awasthi, T. Liang, M.M. Khan, D.P.A. Kilgour, D.R. Goodlett, Y.A. Goo, Comprehensive glycosylation characterization of therapeutic mAbs by top- and middle-down mass spectrometry, Data Brief (2015) (Submitted).
[22] Y.D. Liu, A.M. Goetze, R.B. Bass, G.C. Flynn, N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies, J. Biol. Chem. 286 (2011) 11211–11217, http://dx.doi.org/10.1074/jbc.M110.185041.
[23] W. Xu, Y. Peng, F. Wang, B. Paporello, D. Richardson, H. Liu, Method to convert N-terminal glutamine to pyroglutamate for characterization of recombinant monoclonal antibodies, Anal. Biochem. 436 (2013) 10–12, http://dx.doi.org/10.1016/j.ab.2013.01.017.
[24] J.M. Berg, J.L.S.L. Tymoczko, Section 4.2 Amino Acid Sequences Can Be Determined by Automated Edman Degradation, 5th Ed., BiochemicalW H Free, New York, 2002.
[25] K. Demeure, L. Quinton, V. Gabelica, E. De Pauw, Rational selection of the optimum MALDI matrix for top-down proteomics by in-source decay, Anal. Chem. 79 (2007) 8678–8685, http://dx.doi.org/10.1021/ac070849z.
[26] D. Debois, N. Smargiasso, K. Demeure, D. Asakawa, T.A. Zimmerman, L. Quinton, E. De Pauw, MALDI in-source decay, from sequencing to imaging, Top. Curr. Chem. 331 (2013) 117–141, http://dx.doi.org/10.1007/128-2012-363.
[27] H. Van der Wel, S.Z. Fisher, C.M. West, A bifunctional diglycosyltransferase forms the Fucα1,2Galβ1,3-disaccharide on Skp1 in the cytoplasm of dictyostelium, J. Biol. Chem. 277 (2002) 46527–46534, http://dx.doi.org/10.1074/jbc.M208824200.
[28] P. Chaurand, B.B. DaGue, S. Ma, S. Kasper, R.M. Caprioli, Strain-based sequence variations and structure analysis of murine prostate specific spermine binding protein using mass spectrometry, Biochemistry 40 (2001) 9725–9733, http://dx.doi.org/10.1021/bi010424l.
[29] D. Ayoub, W. Jabs, A. Resemann, W. Evers, C. Evans, L. Main, C. Baessmann, E. Wagner-Rousset, D. Suckau, A. Beck, Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques, MAbs 5 (2013) 699–710, http://dx.doi.org/10.4161/mabs.25423.
[30] K. Srzentic, L. Fornelli, U.A. Laskay, M. Monod, A. Beck, D. Ayoub, Y.O. Tsybin, Advantages of extended bottom-up proteomics using Sap9 for analysis of monoclonal antibodies, Anal. Chem. 86 (2014) 9945–9953.