

Title Page

- Article Title: **The NISTmAb Tryptic Peptide Spectral Library for Monoclonal Antibody**

Characterization

- Journal name: **MABs**
- The full name and details of the corresponding author:

Qian Dong, qian.dong@nist.gov, 3019752569

- The full names and details of all coauthors of the paper:

Yuxue Liang, yuxue.liang@nist.gov, 3019754417

Xinjian Yan, xinjian.yan@nist.gov, 3019753758

Sanford P. Markey, sanford.markey@nist.gov, 3019758696

Yuri A. Mirokhin, yuri.mirokhin@nist.gov, 3019754303

Dmitrii V. Tchekhovskoi, Dmitrii.tchekhovskoi@nist.gov, 3019754673

Tallat H. Bukhari, tallat.bukhari@nist.gov, 3019752516

Stephen E. Stein, stephen.stein@nist.gov, 3019752505

- Acknowledgements

This work was supported solely with NIST funds. Certain commercial instruments or materials are identified in this paper to specify the experimental procedure adequately. Such an identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or instruments identified are necessarily the best available for the purpose.

The NISTmAb tryptic peptide spectral library for monoclonal antibody characterization

Qian Dong*, Yuxue Liang, Xinjian Yan, Sanford P. Markey, Yuri A. Mirokhin, Dmitrii V.

Tchekhovskoi, Tallat H. Bukhari, and Stephen E. Stein

Biomolecular Measurement Division, National Institute of Standards and Technology, 100

Bureau Drive,

Stop 8362, Gaithersburg, Maryland 20899, United States

* Corresponding author: qian.dong@nist.gov

Running title: **The NISTmAb tryptic peptide spectral library**

Abbreviations Page

1D LC-MS, one-dimensional liquid chromatography-mass spectrometry; 2D LC-MS, two-dimensional liquid chromatography-mass spectrometry; CDR, complementarity-determining region; CHO, Chinese hamster ovary; DTT, dithiothreitol; ESI, electrospray ionization; HCD, higher-energy collision dissociation; HPLC, high-performance liquid chromatography; HSA, human serum albumin; Humira®, brand name for the drug adalimumab; IAA, iodoacetamide; LC-MS, liquid chromatography-mass spectrometry; LC-MS/MS, liquid chromatography-tandem mass spectrometry; mAbs, monoclonal antibodies; MS1, full MS scan; MS2, tandem MS scan; MS-GF+, software for peptide identification; MWCO, molecular weight cutoff; NIST, National Institute of Standards and Technology; NISTmAb, reference material 8671, humanized IgG1κ monoclonal antibody; NIST MSQC, a software platform for analysis of shotgun proteomics; NS0, mouse myeloma cell line; PSMs, peptide-spectrum matches; PTM, post-translational modification; RapiGest, brand-name for sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate; RT, retention time; stdev, standard deviation; TCEP - Tris(2-carboxyethyl)phosphine; TFE, 2,2,2-trifluoroethanol; Tris HCl, Tris(hydroxymethyl)aminomethane hydrochloride

Abstract

We describe the creation of a mass spectral library composed of all identifiable spectra derived from the tryptic digest of the NISTmAb IgG1 κ . The library is a unique reference spectral collection developed from over six million peptide-spectrum matches acquired by liquid chromatography-mass spectrometry (LC-MS) over a wide range of collision energy. Conventional one-dimensional (1D) LC-MS was used for various digestion conditions and 20- and 24-fraction two-dimensional (2D) LC-MS studies permitted in-depth analyses of single digests. Computer methods were developed for automated analysis of LC-MS isotopic clusters to determine the attributes for all ions detected in the 1D and 2D studies. The library contains a selection of over 12,600 high-quality tandem spectra of more than 3,300 peptide ions identified and validated by accurate mass, differential elution pattern, and expected peptide classes in peptide map experiments. These include a variety of biologically modified peptide spectra involving glycosylated, oxidized, deamidated, glycated, and N/C-terminal modified peptides, as well as artifacts. A complete glycation profile was obtained for the NISTmAb with spectra for 63% and 100% of all possible glycation sites in the heavy and light chains, respectively. The site-specific quantification of methionine oxidation in the protein is described. The utility of this reference library is demonstrated by the analysis of a commercial monoclonal antibody (adalimumab, Humira®), where 691 peptide ion spectra are identifiable in the constant regions, accounting for 60% coverage for both heavy and light chains. The NIST reference library platform may be used as a tool for facile identification of the primary sequence and post-translational modifications, as well as the recognition of LC-MS method-induced artifacts for human and recombinant IgG antibodies. Its development also provides a general method for creating comprehensive peptide libraries of individual proteins.

Keywords: NISTmAb, tryptic peptide spectral library, monoclonal antibody, high resolution, HCD fragmentation, post-translational modifications (PTM), quantitation, 1D/2D LC-MS/MS, peptide classification

Introduction

Since their first approval by the US Food and Drug Administration in 1986, monoclonal antibodies (mAbs) have emerged as one of the fastest growing classes of protein therapeutics in the treatment of various human diseases.¹⁻² This success has evolved into new generations of products, such as bispecific mAbs, mAb fragments, antibody-drug conjugates, and other derivatives.³⁻⁴ However, because of their large size and the high degree of heterogeneity that may arise from various post-translational and chemical modifications during cell culture, purification, formulation, and storage, the characterization of these proteins continue to present a substantial challenge.⁵⁻⁷ In an effort to support and promote high-quality measurements, the National Institute of Standards and Technology (NIST) issued a humanized IgG1 κ mAb reference material, RM 8671 (NISTmAb).⁸ This reference mAb embodies the quality and characteristics of a donated biopharmaceutical product and has been characterized by NIST researchers, industry and academic community collaborators.⁹ As a part of the NISTmAb project to provide reference data, we report here a high-resolution spectral library of peptides derived from this molecule. Because all IgG antibodies have very similar constant regions, this library should be a useful tool for analyzing these regions for all human IgG and IgG-based biotherapeutics.

Over the past decade, untargeted bottom-up LC-MS has become the method of choice for mAb characterization at the peptide level.¹⁰⁻¹¹ Despite this progress, the ability of an LC-MS analysis to determine all modifications on each residue of mAb therapeutics has not been demonstrated.¹²⁻¹⁴ This approach typically involves the digestion of denatured proteins by trypsin or other proteases, followed by the separation of peptides by LC-MS, and their identification by sequence search methods from tandem mass spectra. Here, we report development of an alternative peptide identification method based on matching experimentally derived spectra in a comprehensive library. Special strategies were developed and applied for identifying all major modifications at each residue level because conventional peptide identification methods have been developed for mixtures of thousands of proteins containing a limited number of targeted modifications. We have previously built a peptide spectral library for human serum albumin (HSA),¹⁵ and, more recently, a glycopeptide library of high-resolution spectra was derived for the NISTmAb.¹⁶⁻¹⁷ These glycopeptides are invisible to most commonly used database identification programs since they fragment primarily by the loss of sugar units and reveal little of the peptide backbone sequence under commonly used collision energy conditions. In that library, a total of 60 different glycans identified from more than 200 different glycopeptides were represented for the NISTmAb. In this work, we extended the prior work in order to construct an extensive library of all identifiable peptides from the NISTmAb. A wide range of digestion conditions were used, as well as more comprehensive in-depth results from 24-fraction 2D LC separations, to acquire as many relevant peptide ion spectra as possible, including many short and long peptides and charge states. We show that, in addition to the known advantages of speed and sensitivity when using a library search, the NISTmAb library-based identification of modified peptides is advantageous over the database search tools for readily and reliably

determining identities of low levels of uncommon modified species in antibody drugs. The detection of these modifications, challenging for search engine analyses, can be achieved without using special settings or procedures.

Results

1. Identification of Library Reference Spectra

1.1 Overview of MS/MS Spectra of the Library

A high-resolution mass spectral library of all identified peptides produced in the tryptic digests of the NISTmAb sample was created for comprehensive analysis of mAbs. This library is composed of over 12,600 high-quality tandem spectra of more than 3,300 peptide ions. The 12 varieties of spectra included in the library are shown in Table 1. The second variety consists of 1,700 spectra of glycopeptides presented in a published paper that can be used together with this library.¹⁶⁻¹⁷ Varying collisional energies of higher-energy collision dissociation (HCD) were used and spectra were generated by 1D LC for 13 different digestion conditions as well as by 2D LC (20- or 24-fraction) studies of single digests (Materials and Methods and Supplementary Material Table S-1). The 1D studies (264 runs) cover a wide range of common digestion conditions, in order to generate the widest possible variety of peptides of analytical concern. No attempts were made to optimize digestion; instead, we employed distinct but complementary protocols to achieve a broad coverage of peptides. A study of optimization of the NISTmAb tryptic digestion has been reported.¹⁸ The 2D studies (553 runs) increased the depth of coverage and quality of these peptide spectra. The library of annotated mass spectra of all NISTmAb peptide ions

acquired in this work, along with the library search software, is freely available at http://www.chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:human_igg1k_mab_drugs.

1.2 Peptide Classification

The number of distinct peptide spectra in the library organized for full antibody chains as well as their constant regions are shown in Table 2. These numbers include distinct library sequences, peptides (sequence + modification), peptide ions (sequence + modification + charge state), spectra (sequence + modification + charge state + collision energy), and the total number of underlying peptide-spectrum matches (PSMs).

This work employed a peptide classification scheme developed earlier for building spectral libraries of tryptic peptides for a single protein, HSA.¹⁵ This scheme divided all peptides into the following six classes.

- Class 1. No missed tryptic-cleavage (includes K/R at the N-terminal resulting from cleavage between adjacent cleavable residues, and K/R followed by proline¹⁹)
- Class 2. Expected missed tryptic-cleavage (D, E, K, or R within 1 residue of the missed cleavage site)
- Class 3. Unexpected missed tryptic-cleavage (not in Class 2)
- Class 4. Missed alkylation on cysteine or alkylation on other amino acids
- Class 5. Modification (a. common, and b. other post-translational or method-induced)
- Class 6. Semi-tryptic peptides with non-specific cleavage at one terminal, including the cleavage of K/R followed by proline (a. in-source and b. in-solution).

As developed in the earlier work,¹⁵ peptides with properties of two or more classes 3, 4, 5b, or 6b are generally rejected as improbable peptide features (see examples in Discussion). In this scheme, multiple unexpected miscleavages are allowed, but multiple, unexpected modifications are not. An ideal digestion profile from an analyst's perspective would be one where all the expected tryptic peptides (Class 1 or Class 2) were abundant, while the others (Class 3, 4, 5, or 6) were not. In the following sections, this classification scheme is applied to nine experiments in which each sample prepared using Protocol 4 was digested for 0.25 h, 2 h, and 18 h, with LC-MS analysis in triplicate (see GuanRT/TCEP²⁰ in **Supplementary Material Table S-1 and Figure S-1**). The comparison of numbers and abundances of peptides (e.g., median values from triplicate experiments of the same digestion) in each class is given in Figure 1.

Class 1 Fully tryptic peptides. These peptides are products of simple tryptic cleavage,¹⁹ and therefore expected to represent the bulk of the ion intensity in tryptic digests of proteins. Other identified peptides often originate from modifications to these peptides. Figure 1 A shows that 93, 118, and 122 peptides of Class 1 were detected in three triplicate analyses at the digestion time points (0.25, 2 and 18 h), respectively. All major fully cleaved tryptic peptide ions were observed at short digestion times, with significantly higher signal intensity for many at longer times, yielding about 39%, 54% and 74% of total ion intensity at 0.25, 2 and 18 h, respectively, as illustrated in Figure 1 B.

Class 2 Peptides with expected missed-cleavage. Tryptic cleavage rates can be reduced by the presence of adjacent cleavage sites or acidic residues.²¹ Depending on the degree of digestion, these can be found at high relative abundances, even though they are not generally targets in peptide mapping studies. At short digestion times (0.25 and 2 h), Class 2 accounts for 9% and 7% of identified peptides (Figure 1 A), and 25% and 21% of total intensity (Figure 1 B), respectively. Approximately 30% of the 20 top-ranking peptides identified belonged to Class 2. Such peptides are frequently observed with trypsin digestion at shorter times.^{15, 22-23} Longer Class 2 peptides can be beneficial in achieving full sequence coverage and can assist in the determination of protein sequence and modification.

Class 3 Peptides with unexpected missed cleavage sites. Class 3 peptides generally result from incomplete digestion, probably because of insufficient denaturation, which left certain peptide bonds less accessible to digestion. As seen from Figure 1 A, Class 3 comprises fewer identifications than Classes 1 and 2 peptides. Intensities of these peptides are low compared to those of Classes 1 and 2 (Figure 1 B), representing 8%, 3% and 3% of total intensities at 0.25 h, 2 h, and 18 h, respectively. Class 3 peptides can be used to assess digestion efficiency of different sample preparation methods.

Class 4 Peptides with under/over alkylation. Class 4 measures undesirable misalkylation of cysteine residues or alkylation of other residues in a protein digest with iodoacetamide.²⁴⁻²⁵ Figure 1 A shows less than 5% of Class 4 peptides relative to all identified peptides. Their summed abundances are typically below 3% of the total intensity identified in our triplicate

experiments. Misalkylated peptides were generally found to elute 1 - 4 min after their alkylated counterpart. It should be noted that about 30 peptides were alkylated at a site other than cysteine residue, the most significant of which are shown in Table 3. Methionine was found to be the most frequent target of carbamidomethylation, as has been previously reported.²⁵ Other overalkylated sites also include serine and the N-terminal amino group.

Class 5 Modifications. Class 5 peptides include biological post-translational modifications (PTM) present in the original sample as well as analytical artifacts produced during liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. The former is often an important target of proteomics, whereas the latter is a nuisance, especially when impeding the analysis of the former.²⁶ As seen in Figure 1 A, between 665 to 744 different modified peptide ions were identified in a triplicate experiment at different digestion times, constituting 49% to 55% of identified product ions. The ion intensity for this class accounted for 15% to 18% of the total identified ion intensity (Figure 1 B), making it the third most abundant class. Among all Class 5 modifications, Met/Trp oxidation, N-terminal Cys or Gln loss of ammonia, and water loss from N-terminal Glu are commonly observed in proteomic experiments, while other modifications are dependent on the target of analytical interest or specific experimental conditions (see *Peptide Classification Scheme* in Materials and Methods). These distinct types of modifications are further divided as subclasses 5a (common) and 5b (rare) in identifying false positives during the library development (see Discussion).

Class 6 Semi-tryptic Peptides. Class 6 peptides have one terminus resulting from a non-tryptic cleavage. From the results of triplicate analyses by three separate tryptic digests at 0.25 h, 2 h and 18 h, a total of 219 (18%), 293 (22%) and 415 (28%) different semi-tryptic peptide ions were detected, respectively (Figure 1A). However, these are low abundance peptides accounting for only 4% to 7% of the total abundance identified. These peptides are separated into two subclasses.

(a) In-source fragments. These are well-known artifacts easily identified by their coelution with their more abundant presumed precursor tryptic peptide. Their precursor m/z must match major fragment ions in the spectrum of the parent ion under lower collisional energies. In our single experiment, typically 30% of all identified semi-tryptic peptides originate from in-source fragmentation. For example, a prominent in-source peptide PAPIEK arose from ALPAPIEK, where the PAPIEK fragment was dominant, consistent with the known preference for N-terminal cleavage of proline and the facile formation of B2-ions from short peptides.

(b) “In-solution” fragments. Peptides arise from irregular cleavage during sample preparation or digestion. These presumably result from inaccurate trypsin cleavage due to incomplete denaturation or related anomalies. Most identified semi-tryptic peptide ions belong to this subclass. Of the total identified semi-tryptic peptides, 157, 210, and 298 semi-tryptic peptides were found for 0.25 h, 2 h and 18 h digestions, respectively, all with elution times different than their potential parent tryptic peptides.

1.3 Primary Sequence Peptides

The representative tryptic peptides derived from the NISTmAb are shown in Table 4, acquired from the same triplicate experiment with 18 h of digestion that was used for reporting peptide classes (see Section 1.2). Most are Class 1 (fully tryptic, unmodified peptides), but a few are Class 5 peptides, representing major protein modifications: N-terminal pyroglutamic acids (H1), N-linked glycopeptides (H22 and H23), and C-terminal lysine clip (H38). Fifty-eight of 62 possible fully-cleaved peptides by trypsin were detected in the 1D LC-MS/MS analyses, and four possible dipeptides were not detected (Table 4). The LC-MS attributes (theoretical precursor m/z , absolute and relative abundance, and retention time (RT)) of each peptide ion were extracted from raw data and determined in their most abundant charge state. The median values of these attributes were calculated across triplicate analysis of the same NISTmAb digest. The mass accuracy of all identified peptides is within 5 ppm.

Early-eluting peptides. Peptides having five or fewer amino acids account for 40% of all theoretical peptides and 14% of the amino acid residues in NISTmAb. Since it is often challenging to find short peptides due to their generally poor resolution of early eluting peptides, they were identified by the targeted analysis like that developed for glycation in the next section. As shown in Table 4, most short peptides eluted in the solvent front along with digest buffer components²⁷ where their signals are reduced due to ion suppression and chromatographic peak broadening (data not shown). This explains both their lower detection rate and variable abundance compared to the better-resolved peptides.

1.4 Modified Peptides with Complex Fragmentation Characteristics

A total of 24 different modification categories were clearly identified (Table 5) and the probable origins of each category are noted in the last column. All modified peptide spectra are included in the library. Among them, some common modifications such as deamidation, oxidation, or formylation can be readily identified from their tandem spectra by mass shifts in precursor and product ions. However, others such as glycosylation, glycation, alkylation of methionine, or metal-ion adduction can greatly influence peptide fragmentation. For these latter modifications, special methods are needed for their identification and are discussed below.

Glycated peptides. Protein glycation is the non-enzymatic adduction of a sugar molecule on the ϵ -amino groups on lysine or on a protein N-terminus.²⁸ It most commonly leads to a mass increase of 162 Da by a hexose sugar. Glycation presents a special challenge because the fragmentation of glycated peptides involves the breakdown of the sugar, generating abundant product ions corresponding to various characteristic neutral losses, but limited peptide sequence ions. The localization of glycation on lysine residues impedes trypsin cleavage and, combined with their low abundance, adds to identification difficulties. Most conventional sequence search engines usually miss these modifications, and in view of their significance in mAb quality measurement, a special workflow involving both 1D and 2D LC-MS/MS analyses directed for their identification was developed. This is shown in Figure 2 and summarized below.

(1) A list was first created of m/z , charge state, and RT of all library-identified, unmodified peptides (Classes 1, 2, and 3) containing either lysine or the protein N-terminus and containing up to three missed cleavages.

(2) For each ion in the above list, all precursor ions with masses greater by a hexose unit (within 5 ppm) and within a 10-min retention window were marked as possible glycosylated peptides.

(3) Identifications were validated by ensuring a constant shift in RT for a given glycosylated peptide ion relative to its non-glycosylated analog. It was found that virtually all glycosylated peptides eluted within 3 min of their unmodified analog. Consequently, spectra of components eluting outside this range were rejected.

(4) Candidate spectra obtained in Step 3 were confirmed if all major fragment ions were consistent with the peptide and they contained no major unexplained fragments (>20% of the base peak). Specifically, various characteristic neutral losses were associated with glycosylation, four of which were frequently observed: 3 H₂O (54.03 Da), 3 H₂O + HCHO (84.04 Da), 4 H₂O (72.04 Da), and C₆H₁₀O₅ (162.05 Da), from the precursor, y, b ion series. Filters in the Steps 3 and 4 led to rejection of approximately 10% of the candidate spectra.

Using the above workflow, a total of 276 spectra of 58 glycosylated peptide ions were acquired by 1D LC-MS/MS analysis. Additionally, 590 spectra were identified for 92 glycosylated peptide ions in the 2D LC-MS/MS fractionation runs, covering a range of different collisional energies. Table 6 summarizes the results of the glycosylated peptides identified in at least two runs of the 1D or 2D

studies. These data include peptide sequence, charge states, ion intensity (total peak area summed over charge states), as well as the observed RT (min) and the RT difference between glycosylated and non-glycosylated peptides. All observed glycosylated peptides and their relative RT in the 1D analyses were confirmed in the 2D analyses.

The median RT values of the high-performance liquid chromatography (HPLC) for glycosylated peptides and their RT difference from that of the unmodified forms in Table 6 reveal that most of glycosylated peptides exhibit a consistent, small RT shift within ± 1 min from their unmodified forms. The exceptions are several short peptides that are shifted by as much as 3 min. Figure 3 shows the retention shifts of glycosylated peptides. Generally, glycosylated peptides elute after their unmodified forms when RT < 40 min and before their unmodified forms when RT > 40 min.

The 2D separations also found several unexpected glycosylation sites. As noted in Table 6, glycosylation was found at the C-terminal lysine residue in five peptides and at the N-terminus of three peptides. These identifications were carefully confirmed and validated by manual inspection of the data from full MS scan (MS1) and tandem spectra (MS2). Examples are included in Supplementary Material Figure S-2. We suspect that these peptides may have been glycosylated during sample processing.

Glycosylation is known to alter peptide fragmentation due to the ease of fragmenting the sugar group, generally resulting in few sequence-specific product ions, thereby hindering their automated identification. This is illustrated in Figure 4, which compares spectra of a glycosylated to

a non-glycated peptide. As shown here, the mass spectrum of a glycated peptide is typically dominated by characteristic, abundant fragment ions corresponding to the losses of water molecules from the labile hexose sugar. All high-quality glycation spectra were annotated and added to the spectral library.

Metallated peptides. Peptides containing metal ion adducts are common low abundance artifacts in electrospray ionization (ESI).²⁹⁻³⁰ A total of 191 such spectra with sodium, calcium, or iron were identified by MS-GF+,³¹ half of them with low scores. The most common such peptide ion adduct was sodium; 59 sodiated ions were identified. Examination of adduct fragmentation behavior showed that the metal atom did not generally appear to be localized at specific amino acids as suggested by Unimod.³² Rather, metal atoms appear to be bound to multiple amino acids in a peptide sequence, effectively impeding fragmentation in the relevant region, and reducing their “scores” from sequence search methods. Furthermore, the formation of a unique characteristic product ion was frequently observed in singly-charged sodiated peptides via rearrangement involving the loss of the C-terminal residue upon fragmentation.

Fragmentation characteristics of sodiated peptides are examined and compared to their non-sodiated counterparts in Figure 5. The spectrum without sodium in Figure 5 B exhibits uniform dissociation, yielding both b and y product ion series. Many of these fragment ions are absent in the sodiated spectrum in Figure 5 A. The sodiated spectrum primarily contains abundant singly-charged, sodium-containing y fragment ions: y₇ to y₉ and b₁₀, and low abundance doubly-charged, sodium-containing y ions, y₉ to y₁₂. As shown in Figure 5 A, the impeded fragmentation

region (indicated by a red line) in the sodiated peptide is consistent with sodium bound to these multiple amino acid residues. Note also the fragment at m/z 758.3365, corresponding to a truncated, 2+ ion VGYMHWYQQKPG, apparently arising from a rearrangement involving the loss of the C-terminal Lys, a known diagnostic ion for sodiated peptides.³⁰

Because the fragmentation of metal adducts typically produces less sequence information than corresponding non-adducted peptide ions, their confident identification required confirmation by their co-elution with the non-metallated precursor. Additionally, the peak signal-to-noise ratio requires that there be at least 15% of the adduct-containing peaks with intensities above 10% of the base peak.

Alkylated methionine. As illustrated in Figure 6, the product ions in HCD from these methionine-alkylated peptides show significant non-sequence specific cleavage, with an abundant neutral loss of C_3H_7NOS , making them hard to identify by sequence search engines. The identification of these peptides was conducted using a workflow similar to that developed for glycation summarized in Figure 2 and four steps of the previous ***Glycated peptides*** section. Specifically, all library-identified, unmodified peptide ions (Classes 1 and 2) were collected in the first step, and they were considered as possible overalkylated peptides in Step 2 if their masses are greater by carbamidomethylation (within 5 ppm). The HCD fragmentation analysis included an abundant neutral loss of C_3H_7NOS for methionine alkylation.

2. Applications of the NISTmAb Spectral Library to Structure and PTM Characterization

2.1 Confirmation of NISTmAb Primary Sequence

Peptides in the library cover 99% of the NISTmAb sequence, representing 211 of 213 light chain residues and 444 of 450 heavy chain residues. Though most PTM species of IgG1-based antibodies have very low abundances, three of them have relative intense signals, namely N-terminal pyro-glu, clipped C-terminal Lys, and N-linked glycosylation (Table 4). As reported previously, the N-linked glycosylation site was approximately 99% occupied.¹⁶ The N-terminal glutamine residue was almost completely converted to pyroglutamic acids and approximately 90% of the terminal Lys was truncated. These observations obtained in the present 1D study agree with previous NISTmAb analysis reports.³³

2.2 Determination of Glycation Sites on NISTmAb

Glycation of mAbs can occur at low levels during biomanufacturing where sugars are used as energy and carbon sources. In principle, glycation could occur on every available free amino site across the entire recombinant antibody. However, only a limited number of glycated sites detected on the heavy and light chains have been reported so far using standard analytical approaches.³⁴⁻³⁷ Their low abundance and interference with tryptic cleavage have made these modifications difficult to identify and quantify.

NISTmAb has a total of 51 potential glycation sites, with an N-terminal amine for each antibody chain, and 35 and 14 lysine residues in the heavy chain and light chain, respectively. From the present 1D LC-MS/MS runs, glycation was found on 11 heavy and 8 light chain lysine residues. From 2D LC-MS/MS runs, a total of 22 (63% of possible) glycation sites were detected in the

heavy chain and 15 (100%) in the light chain, or nearly 3/4 of all possible sites. These glycation sites detected in the present 1D study are generally in agreement with previous reports,³⁵⁻³⁶ while additional 16 glycosylated sites are detected by the 2D study.

Potential glycation sites on lysine residues are distributed across the entire NISTmAb as shown in Columns 1 and 2 of Table 6. There are 13 in the constant domains (CH₁, CH₂, and CH₃) of the heavy chain and 9 in the constant region of the light chain, all of which should be applicable to other IgG1 antibodies. Glycation at lysine residues 58 and 66 in the heavy chain and lysine 52 in the light chain are involved in the complementarity-determining region (CDR) 2, and thus they could be critical sites to monitor due to the known biological function in the CDR regions for antigen binding. The C-terminal lysine of the heavy chain was also glycosylated at a low level. Additionally, glycation was observed on the free N-terminus of the light chain in 2D LC-MS/MS runs, although the level was very low compared to that of lysine residues, consistent with the observation by Gadgil et al.³⁸ No glycation was detected on the heavy chain N-terminus by either 1D or 2D runs, probably due to the conversion of nearly all terminal glutamic acid residue to pyroglutamic acid.

To summarize, glycosylated peptide spectra were identified for most lysine residues in the NISTmAb. The light chain potential sites are documented for the first time to be fully accessible to glycation. However, all identified glycosylated peptides are minor species, with the relatively intense glycosylated ions typically having around 0.3% of the intensity of the most abundant ion in the same 1D LC-MS run.

2.3 Quantification of Methionine Oxidation in NISTmAb

We found that the peptides containing oxidized methionine showed complex elution behavior. This has been reported to complicate accurate quantification of site-specific methionine oxidation in antibodies.³⁹⁻⁴¹ To assess the problem in the present work, all eight methionine residues in the NISTmAb were examined. Six of the Met residues are part of the heavy chain with three in the Fv domain (Met 34, Met 87, and Met 101) and three in the Fc domain (Met 255, Met 361, and Met 431), and the other two are part of the light chain Fab domain (Met 4 and Met 32). A more detailed discussion of these findings follows.

Observation of in-sample, in-column and in-source oxidation. We observed that methionine oxidation often produced two closely eluting chromatographic peaks connected by an ion signal of lower and variable abundance. The three selected ion chromatograms of oxidized peptides in Figure 7 are examples that illustrate this “bridge-like” feature. As can be seen in a1, b1, and c1 of Figure 7, unoxidized peptides elute at 31.3, 35.9, and 44.5 min. All oxidation artifacts generated by in-source oxidation coelute with their parent peptides shown in a2, b2, and c2. The in-sample oxidized peptides elute earlier than their unoxidized counterparts. Note the significant continuous signal between the in-sample oxidized and in-source oxidized peaks. The MS1 peaks shown in Figure 7 exhibit similar patterns along this “bridge”, and MS2 spectra from this retention interval were further identified to contain oxidized methionine residues. We postulate that the signal observed between in-solution and in-source oxidized chromatographic peaks is generated by in-column oxidation of unoxidized peptides containing Met residue during their

passage through the column. The origin of the signal between in-sample and in-source peaks was confirmed by data from 2D studies where native and oxidized peptides were separated by the first LC into different fractions (data not shown). In summary, peptides containing a methionine residue are susceptible to oxidation in the column leading to the generation of complex chromatographic peaks from pre-column, in-column, and in-source (post-column) oxidation.

Site specific quantification. Oxidized forms of all eight methionine residues were identified and spectra for 298 peptide ions are included in the current library. We inspected four 1D LC-MS analyses, performed on the same day that demonstrated minimal oxidation artifacts, to manually separate the effects of in-sample from in-column and in-source oxidation and quantify the relative percentage of all oxidized methionine residues. The resulting levels (as well as standard deviations) of in-sample oxidation in the peptides containing Met sulfoxide are in good agreement with the similar manual analysis reported by an interlaboratory study on identification and quantification of the NISTmAb methionine oxidation (Figure 8).³⁶ Consistent with that study, Met 255 and Met 431 are the two most oxidized sites, with 3.4% and 3.0% abundance relative to the corresponding unoxidized sites, respectively. Next are Met 34 of the heavy chain and Met 32 of the light chain, with approximately 2% oxidation. Met 361, Met 87 and Met 4 exhibit low levels of observed oxidation, at 0.7%, 1.3%, and 1.4%, respectively. As demonstrated in both studies of the NISTmAb,³⁶ accurate methionine quantification requires careful separation of native and artifact oxidized peptides.

Discussion

False positive identifications. The NISTmAb spectral library is intended to contain all identifiable peptides derived from a tryptic digest. Because of the wide variety of peptides and modifications, and their wide distribution of relative abundances, it is not possible to derive statistically meaningful false discovery rates, as is widespread practice in global proteomics where thousands of proteins containing a very limited number of modifications are analyzed.^{15, 42} In this work, all identifications were made using sequence engines, targeted analyses and a variety of peptide classification rules. One of the most significant rules is not allowing more than two rare modifications/cleavages on a single peptide (e.g., two iron atoms in a single peptide). Rare features include unexpected miscleavages, under- or over-alkylation, uncommon and low abundance modifications, and semi-tryptic peptides (Classes 3, 4, 5b, and 6b), and were generally considered as improbable peptide features. Practically, as a rule, while the assignment of two or more rare modifications would trigger spectrum rejection, multiple missed cleaved or alkylated sites were allowed as they were frequently observed.

We illustrate two examples of spectra rejected using these rules. The first rejected identification is shown in Figure 9 A. In this example, a spectrum was initially identified as non-alkylated formyl^lSCD^{hex}KTHTCPPCPAPPELLGGPSVFLFPPKPK (heavy chain 222-251) at m/z 839.1618, which is within 4 ppm of its theoretical value at charge state 4+. This false identification exhibits a combination of uncommon features: 1) misalkylation of all three cysteine residues, 2) N-terminal formylation, and 3) glycation of lysine. In other respects, it satisfied criteria for acceptance. We determined that this case might be better explained by the unintended sampling of the ¹³C isotope of a tryptic peptide, specifically the mono-oxidized heavy chain 222-251 peptide. Reexamination of the raw data justified the latter explanation. The second example is

the rejected false identification of the triply-charged, modified light chain N-terminal peptide, as $D^{\text{dehydrated}}I^{\text{cation:2Ca[III]}}QMTQSPSTLSASVGDR$ at m/z 650.933 (Figure 9 B). Although 63% of total product ion abundance was assignable to the dominant y series ions, two postulated N-terminal modifications led to the exclusion of this identification. The more probable assignment resulted from a manual inspection, which found that 75% of the total number of peaks (ranging from 2% to 23% of the base peak) corresponded to b series ions, arising from the loss of C_3H_7NOS from the methionine-alkylated light chain L1 peptide at m/z 650.645. It was also confirmed by MS1 analysis of the ion isotopic m/z values.

In another test of the false positive discriminating potential of this library, we used the NISTmAb library to match spectra from an extensive global proteomic tryptic Jurkat cell digest. As an immortalized line of human T lymphocyte cells, Jurkat cells are not expected to express antibodies. The library search was performed with the NIST MSQC pipeline,⁴³ a software platform for analysis of shotgun proteomics, and identification score threshold was set to 450 (1000 as highest). This search did not yield any identifications above this threshold. In contrast, other NIST human libraries⁴³ identified 25,000 peptide ions in this dataset, an indication of the specificity of the NISTmAb library. Clearly, for the present library, false positive identifications from proteins unrelated to an IgG sample appear very unlikely.

Use of the NISTmAb spectral library in general antibody analysis. Because the NISTmAb constant regions are very similar for all IgG antibodies, the present library is suitable for identifying peptides and modifications originating from these regions. In Figure 10, we

demonstrate this using LC-MS/MS results for Humira®, a mAb drug commercially produced in CHO cells rather than a mouse cell line, which was used for production of the NISTmAb. Because the principal glycosylation site is in the conserved region, the previously published NISTmAb glycopeptide library¹⁷ is directly applicable to all such IgGs (N.B. the glycopeptide library includes sialylated glycans containing Neu5Ac and Neu5Gc). This analysis was carried out using the NIST MSQC Pipeline.⁴³ It identified 691 peptides using a match score threshold of 450, of which 497 are from the heavy chain constant regions, and 194 from the light chain constant region. These include 165 major tryptic peptides (Class 1 and 2), 80 glycopeptides representing 25 different glycan structures (including three sialylated glycans with Neu5Ac), 83 oxidized peptides containing oxidation on Met, Trp, His, and Pro, 15 glycosylated peptides identifying 10 main glycosylation sites, and 141 deamidated peptides from Asn or Gln. Also, shown in gray, dark and light blue are other peptides, which included many analytical artifacts such as sodium adducts and formylation, and semi-tryptic and miscleaved peptides. This example demonstrates the facile re-identification of major peptide classes and known PTMs by a NISTmAb library search. We also showed that artifact peptides can be recognized that might otherwise lead to false positive identifications, and that their identities can be used for assessing the quality of the sample preparation process.

Materials and Methods

Materials

The Sample 8670 NISTmAb (an interim material), lot 3f1b, is an in-house IgG1κ mAb derived from a separate production lot of NISTmAb 8671.⁸ It was expressed by NS0, a model cell line

from murine myeloma, obtained from the Bioanalytical Science Group at NIST. Digestion reagents guanidine hydrochloride, urea, dithiothreitol (DTT), iodoacetamide (IAA) and Sigma trypsin T1426 were purchased from Sigma- Aldrich (St. Louis, MO). Sequencing-grade trypsin was purchased from Promega (Madison, WI). RapiGest, brand-name for sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate, was purchased from Waters (Milford, MA); Zeba spin columns (7K molecular weight cutoff (MWCO)) were purchased from Thermo Fisher Scientific (Waltham, MA). Chromatographic separations were performed on an Acclaim pepmap100 nano column (150 mm × 75 μm, C18, 3 μm particle size, 100 Å pore size, Dionex, Sunnyvale, CA). Humira pen was purchased from Abbott Laboratories (North Chicago, IL).

Tryptic Digestion

General procedure. 500 μg of NISTmAb was denatured using various chemicals and solvents in different conditions in 50 μL of 100 mmol/L Tris HCl (Tris(hydroxymethyl)aminomethane hydrochloride) buffer at either room temperature or high temperature of 85 °C (HT). Reduction was performed by adding 2.5 μL of 200 mmol/L DTT to the above denatured mixture and incubating at room temperature for 1 h, followed by alkylation with 10 μL of 200 mmol/L IAA at room temperature in the dark for 1 h. Next, the excess IAA was quenched by adding 10 μL of 200 mmol/L DTT solution and incubated at room temperature for 1 h. The resulting mixture was subjected to digestion by sequencing-grade trypsin at 37°C for 15 min, 2 h and 18 h for the 1D studies. Samples for 2D studies were denatured in guanidine at high temperature (85 °C) or room temperature for 10 min prior to reduction and alkylation, and digested for 2 h and 18 h. The reaction was quenched by adding 50% (vol/vol) formic acid.

To generate a wide variety of tryptic peptides, parameters for 13 protocol variations were used during denaturing, reduction, and alkylation of the NISTmAb. Each sample was digested with trypsin for 0.25 h, 2h, and 18 h. These were summarized here and in Supplementary Material Table S-1. (1) *UreaRT/DTT*: 6 M urea in Tris HCl buffer was used to denature the sample at room temperature. The alkylated protein solution was subjected to desalting by Zeba spin column (7K MWCO) prior to adding 10 µg of Promega trypsin. (2) *GuanRT/DTT*: 6 M guanidine hydrochloride in Tris HCl buffer was used to denature the sample at room temperature. The alkylated protein solution was subjected to desalting by Zeba spin column (7K MWCO) prior to adding 10 µg of Promega trypsin. (3) *GuanRT/No2DTT*: 6 M guanidine hydrochloride in Tris HCl buffer was used to denature the sample at room temperature without quenching the excess of IAA using the second DTT. The alkylated protein solution was subjected to desalting by Zeba spin column (7K MWCO) prior to adding 10 µg of Promega trypsin. (4) *GuanRT/TCEP*: 6 M guanidine hydrochloride in Tris HCl buffer was used to denature the sample at room temperature. Tris(2-carboxyethyl)phosphine (TCEP) was used as a reducing reagent instead of DTT. The alkylated protein solution was subjected to desalting by Zeba spin column (7K MWCO) prior to adding 10 µg of Promega trypsin. (5) *GuanHT/DTT*: The procedure is same as Protocol 2 with the exception that the denature temperature was set to 85 °C. (6) *HT/DTT*: The sample in Tris HCl buffer was denatured by heating at 85 °C for 1h. 10 µg of Promega trypsin was used to digest the protein. (7) *RapiGestHT/DTT*: 0.2% RapiGest (by volume) in Tris HCl buffer was used to denature the sample at 85 °C. 10 µg of Promega trypsin was used to digest the protein. (8) *RapiGestRT/DTT*: 0.2% RapiGest (by volume) in Tris HCl buffer was used to denature the sample at room temperature. 10 µg of Promega trypsin was used

to digest the protein. (9) *MeOHRT/DTT*: 20% methanol (by volume) in Tris HCl buffer was used to denature the sample at room temperature. 10 µg of Promega trypsin was used to digest the protein. (10) *MeCNRT/DTT*: 20% acetonitrile in Tris HCl buffer was used to denature the sample at room temperature. 10 µg of Promega trypsin was used to digest the protein. (11) *TfeRT/DTT*: 50% TFE (2,2,2-trifluoroethanol) in Tris HCl buffer was used to denature the sample at room temperature. 10 µg of Promega trypsin was used to digest the protein. (12) *GuanRT/DTT/CheapTrypsin10*: The same procedure as Protocol 2 with the exception that 10 µg of Sigma T1426 trypsin was used to digest the protein. (13) *GuanRT/DTT/CheapTrypsin50*: The same procedure as Protocol 12 with the exception that 50 µg of Sigma T1426 trypsin was used to digest the protein.

LC-MS/MS Analysis

ID analysis: The above digests (0.2 µg) were analyzed on a Dionex Ultimate 3000 RSLC (rapid separation liquid chromatography) Nano LC with an Acclaim pepmap100 column with a nanospray source connected to a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) in the positive ion mode. The experimental conditions used were the same as previously reported.¹⁶

2D Analysis: The first dimension of 2D LC analysis was performed by fractionating 1 mg of digests via basic (pH 10) reverse-phase HPLC on an Agilent ZORBAX capillary column with a Thermo Dionex Ultimate 3000 HPLC system. The second dimension of 2D LC-MS/MS analysis was conducted after combining fractions as previously described.¹⁶

Initial Peptide Identifications

Initial identifications were made from HCD spectra derived from tryptic digests using the MS-GF+ search engine³¹ against a fasta file containing the NISTmAb sequence (see Supplementary Material Document S-1). A precursor m/z tolerance of 20 ppm was used and product ion tolerances set by MS-GF+ for HCD fragmentation. In order to reliably identify both long, highly-charged peptides containing multiple missed cleavages as well as peptides containing a wide range of modifications, including semi-tryptic peptides, two separate search engine setting protocols were used. Otherwise, it was found that high scoring, false candidate semi-tryptic peptides with unusual modifications could suppress correct, lower scoring assignments of conventional tryptic peptides. The first search protocol allowed only tryptic peptides and included a list of 24 IgG and method-induced modification categories (see Table 5 and Section 1.4). The second protocol allowed semi-tryptic peptides and permitted only five common modifications (variable cysteine alkylation, methionine oxidation, ammonia loss of N-terminal Gln and Carbamidomethyl-Cys, and water loss from N-terminal Glu). Results of these two searches were merged and identifications with false discovery rate < 1% were collected. Spectra of tryptic peptides with unassigned total abundance greater than 30% were rejected, while this threshold was reduced to 20% for semi-tryptic peptides. In case of multiple identifications of a single peptide ion, the spectrum with the best score for each HCD energy was selected, annotated and added provisionally to the library for later stringent analysis. Note that MS-GF+ did not limit the number of allowed missed cleavages.

Data Analysis

MS1 information for all detectable ions. The detailed analysis of peptide digestion products was done with an in-house program for processing MS1 information (ProMS).¹⁵ The program analyzes isotope clusters of each observed species to determine charge states, RT, monoisotopic m/z , and signal intensity (peak areas derived from all observable isotope peaks) for each ion detected in all LC-MS/MS runs.

Peptide classification scheme. Peptides are separated into six classes (as discussed in *Section 1.2 Peptide Classification*) and fall into two broad categories: common and rare. This is done with the intention of rejecting peptides that contain two or more “rare” features. Common peptides are those expected from digestion and most frequently sought in sequence identification searching, such as fully tryptic peptides or tryptic peptides with unexceptional missed cleavages (near acidic groups or a terminus). These also include the other varieties of peptides found in peptide mapping experiments, for example, Met/Trp oxidation, water loss from N-terminal Glu, and N-terminal Cys or Gln loss of ammonia (commonly observed modifications), as well as in-source products formed during electrospray ionization. Other less commonly observed peptide groups (e.g., missed/over alkylation, miscleaved peptides, semi-tryptic peptides, or uncommon modifications) are classified as rare. The presence of combined rare classes or two or more uncommon characteristics associated with a single peptide would trigger the rejection of the identification. However, frequent and repeat identifications of such peptide ions are manually examined to ensure that no unexpected peptides are lost in the analysis.

Reference

1. Ecker DM, Jones SD, Levine HL. The therapeutic monoclonal antibody market. *MAbs* 2015;7:(1)9-14; PMID: 25529996; <https://doi.org/10.4161/19420862.2015.989042>
2. Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC. Monoclonal antibody successes in the clinic. *Nat Biotechnol* 2005;23:(9)1073-8; doi:10.1038/nbt0905-1073
3. Steiner M, Neri D. Antibody-radionuclide conjugates for cancer therapy: historical considerations and new trends. *Clin Cancer Res* 2011;17:(20)6406-16; doi: 10.1158/1078-0432.CCR-11-0483
4. Reichert JM, Dhimolea E. The future of antibodies as cancer drugs. *Drug Discov Today* 2012;17 (17-18), 954-63; <https://doi.org/10.1016/j.drudis.2012.04.006>
5. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: an update. *Pharm Res.* 2010 27 (4), 544-75; doi: 10.1007/s11095-009-0045-6
6. Berkowitz SA, Engen JR, Mazzeo JR, Jones GB. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat Rev Drug Discov* 2012;11(7)527-40; doi:10.1038/nrd3746
7. Kaltashov IA, Bobst CE, Abzalimov RR, Wang G, Baykal B, Wang S. Advances and challenges in analytical characterization of biotechnology products: mass spectrometry-based approaches to study properties and behavior of protein therapeutics. *Biotechnol Adv* 2012;30 (1):210-22; <https://doi.org/10.1016/j.biotechadv.2011.05.006>

8. National Institute of Standards and Technology. RM 8671 - NISTmAb, Humanized IgG1κ Monoclonal Antibody. Washington (DC): United States Department of Commerce; 2016 Oct 6. https://www-s.nist.gov/srmors/view_detail.cfm?srm=8671
9. Schiel JE, Mire-Sluis A, Davis DL. Monoclonal Antibody Therapeutics: The Need for Biopharmaceutical Reference Materials, State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 1. Monoclonal Antibody Therapeutics: Structure, Function, and Regulatory Space. ACS Symposium Series; Vol. 1176, January 1, **2014**. ISBN13: 9780841230262
10. Leurs U, Mistarz UH, Rand KD. Getting to the core of protein pharmaceuticals-- Comprehensive structure analysis by mass spectrometry. *Eur J Pharm Biopharm* 2015; 93:95-109; PMID: 25791210; <https://doi.org/10.1016/j.ejpb.2015.03.012>
11. Fekete S, Guillarme D, Sandra P, Sandra K. Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Anal Chem*. 2016, 88 (1), 480-507; PMID: 26629607; DOI: 10.1021/acs.analchem.5b04561
12. Zhang ZQ. Large-Scale Identification and Quantification of Covalent Modifications in Therapeutic Proteins. *Anal. Chem*. 2009, 81, 8354–8364; PMID: 19764700; DOI: 10.1021/ac901193n
13. Wang Y, Li X, Liu YH, Richardson D, Li H, Shameem M, Yang X. Simultaneous monitoring of oxidation, deamidation, isomerization, and glycosylation of monoclonal antibodies by liquid chromatography-mass spectrometry method with ultrafast tryptic digestion. *MAbs*. 2016,8 (8), 1477-1486; PMID: 27598507; DOI: 10.1080/19420862.2016.1226715

14. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *The annals of applied statistics*. 2010, 1797–1823; doi: 10.1214/10-AOAS341
15. Dong Q, Yan X, Kilpatrick LE, Liang Y, Mirokhin YA, Roth JS, Rudnick PA, Stein SE. Tandem mass spectral libraries of peptides in digests of individual proteins: Human Serum Albumin (HSA). *Mol. Cell. Proteomics* 2014, 13 (9), 2435–49; doi: 10.1074/mcp.O113.037135
16. Dong Q, Yan X, Liang Y, and Stein SE. In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC-MS/MS. *J Proteome Res*. 2016, 15 (5),1472-86; doi: 10.1021/acs.jproteome.5b01046
17. National Institute of Standards and Technology. Glycopeptide Mass Spectral Library (HCD) of Human IgG1 mAb Drugs. Washington (DC): United States Department of Commerce; 2016 Mar 16. <http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:newigghsalib>
18. Mouchahoir T, Schiel JE. Development of an LC-MS/MS Peptide Mapping Protocol for the NISTmAb. *Anal. Bioanal. Chem.* 2018; in press.
19. Barrett AJ, Rawlings ND, Woessner J. *Handbook of Proteolytic Enzymes* 3rd Edition. Academic Press, London. 3rd December 2012. ISBN13 9780123822192
20. Getz EB, Xiao M, Chakrabarty T, Cooke R, Selvin PR. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem.* 1999, 273 (1), 73-80; PMID: 10452801; DOI: 10.1006/abio.1999.4203
21. Siepen JA, Keevil EJ, Knight D, and Hubbard SJ. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J. Proteome Res.* 2007, 6, 399–408; PMID: 17203985; DOI: 10.1021/pr060507u

22. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 2014, 42, D503–509; <http://doi.org/10.1093/nar/gkt953>
23. Yen CY, Russell S, Mendoza AM, Meyer-Arendt K, Sun S, Cios KJ, Ahn NG, and Resing KA. Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra. *Anal Chem.* 2006, 78 (4), 1071-84; PMID: 16478097; DOI: 10.1021/ac051127f
24. Sechi S, and Chait BT. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* 1998, 70, 5150–5158; PMID: 9868912; DOI: 10.1021/ac9806005
25. Boja ES, and Fales HM. Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* 2001, 73, 3576–3582; PMID: 11510821; DOI: 10.1021/ac0103423
26. Keller BO, Sui J, Young AB, Whittall RM. Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta.* 2008, 627 (1), 71-81; PMID: 18790129; DOI: 10.1016/j.aca.2008.04.043
27. Julka S, Regnier, FE. Benzoyl derivatization as a method to improve retention of hydrophilic peptides in tryptic peptide mapping. *Anal Chem.* 2004 Oct 1;76(19):5799-806; PMID: 15456300; DOI: 10.1021/ac049688e
28. Wei B, Berning K, Quan C, Zhang YT. Glycation of antibodies: Modification, methods and potential effects on biological functions. *MAbs*, 2017, 9 4, 586-594; PMID: 28272973; DOI: 10.1080/19420862.2017.1300214

29. Newton KA, McLuckey SA. Generation and manipulation of sodium cationized peptides in the gas phase. *Journal of the American Society for Mass Spectrometry*, 15 (4), 607-615; PMID: 15047065; DOI: 10.1016/j.jasms.2003.12.014
30. Rožman M, Gaskell SJ. Non-covalent interactions of alkali metal cations with singly charged tryptic peptides. *J Mass Spectrom*. 2010, 45 (12), 1409-15; PMID: 21031360; DOI: 10.1002/jms.1856
31. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun*. 2014,5, 5277; doi:10.1038/ncomms6277
32. Creasy DM, and Cottrell JS. Unimod: Protein modifications for mass spectrometry. *Proteomics* 2004; 4:1534–1536; PMID: 15174123; doi:10.1002/pmic.200300744
33. Formolo T, Ly M, Levy M, Kilpatrick L, Lute S, Phinney K, Marzilli L, Brorson K, Boyne M, Davis D, Schiel J. Determination of the NISTmAb Primary Structure. *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 2. Biopharmaceutical Characterization: The NISTmAb Case Study, Chapter 1*, 2015, pp 1-62, DOI: 10.1021/bk-2015-1201.ch001
34. Saleem RA, Affholter BR, Deng S, Campbell PC, Matthies K, Eakin CM, Wallace A. A chemical and computational approach to comprehensive glycation characterization on antibodies. *MAbs*. 2015, 7 (4), 719-31; PMID: 26030340; DOI: 10.1080/19420862.2015.1046663
35. Zhang J, Zhang T, Jiang L, Hewitt D, Huang Y, Kao YH, Katta V. Rapid identification of low level glycation sites in recombinant antibodies by isotopic labeling with ¹³C₆-reducing sugars. *Anal Chem*. 2012 84 (5), 2313-20. doi: 10.1021/ac202995x.

36. Li W, Kerwin JL, Schiel J, Formolo T, Davis D, Mahan A, Benchaar SA. Structural Elucidation of Post-Translational Modifications in Monoclonal Antibodies. State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 2. Biopharmaceutical Characterization: The NISTmAb Case Study. January 1, 2015, 119-183, DOI:10.1021/bk-2015-1201.ch003
37. Habegger M, Bomans K, Diepold K, Hook M, Gassner J, Schlothauer T, Zwick A, Spick C, Kepert JF, Hienz B, et al. Assessment of chemical modifications of sites in the CDRs of recombinant antibodies: Susceptibility vs. functionality of critical quality attributes. *MAbs*. 2014, 6 (2), 327-39; doi: 10.4161/mabs.27876
38. Gadgil HS, Bondarenko PV, Pipes G, Rehder D, McAuley A, Perico N, Dillon T, Ricci M, Treuheit M. The LC/MS analysis of glycation of IgG molecules in sucrose containing formulations. *J Pharm Sci*. 2007, 96 (10), 2607-21; PMID: 17621682; DOI: 10.1002/jps.20966
39. Zang L, Carlage T, Murphy D, Frenkel R, Bryngelson P, Madsen M, Lyubarskaya Y. Residual metals cause variability in methionine oxidation measurements in protein pharmaceuticals using LC-UV/MS peptide mapping. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2012, 895-896, 71-6; PMID: 22483985; DOI: 10.1016/j.jchromb.2012.03.016
40. Steen H, Mann M. Similarity between condensed phase and gas phase chemistry: fragmentation of peptides containing oxidized cysteine residues and its implications for proteomics. *J Am Soc. Mass Spectrom*. 2001, 12 (2), 228-32; PMID: 11212008; DOI: 10.1016/S1044-0305(00)00219-1

41. Trapp O, Bremer S, Weber SK. Accessing reaction rate constants in on-column reaction chromatography: an extended unified equation for reaction educts and products with different response factors. *Anal Bioanal Chem.* 2009, 395 (6), 1673-9; PMID: 19669676; DOI: 10.1007/s00216-009-2993-4
42. Sennels L, Bukowski-Wills JC, Rappsilber J. Improved results in proteomics by use of local and peptide-class specific false discovery rates. *BMC Bioinformatics.* 2009, 10, 179; PMID: 19523214; DOI: 10.1186/1471-2105-10-179
43. National Institute of Standards and Technology. The NIST Libraries of Peptide Tandem Mass Spectra and Available Windows Software for Peptide Mass Spectral Libraries. Washington (DC): United States Department of Commerce; 2016 Aug 8. <http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:cdownload>, and <http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:msqcpipeline>.

Supporting Information

Supplementary Material Document S-1. NISTmAb Sequence Summary

Supplementary Material Table S-1. Parameters for 13 protocol variations used during denaturing, reduction, and alkylation of the NISTmAb. Each sample was digested with trypsin for 0.25 h, 2 h and 18 h

Supplementary Material Figure S-1. Comparison of ten tryptic digestion protocols using peptide class analysis

Supplementary Material Figure S-2. MS1 and MS2 spectra for an example of the new-found glycosylated peptide, ALPAPIEK, containing glycosylation of the C-terminal lysine.

Figure Legends

Figure 1. Summed peptide identifications and abundances in six classes obtained by median results of triplicate analyses from three separate tryptic digests at (0.25, 2 and 18) h. In all digests, the sample was denatured by guanidine at room temperature and reduced by TCEP (GuanRT/TCEP).

Figure 2. Workflow for identifying peptides and sites of glycation in the NISTmAb.

Figure 3. The LC-MS elution characteristics of glycated versus non-glycated peptides. Red dots indicate glycated peptides eluting after the unmodified form (most with fewer than 10 residues) while blue dots eluting before the unmodified form.

Figure 4. Tandem mass spectra of (A) triply-charged glycated peptide ion at 21 eV, DSTYLSSTLTLKADYEK, compared with (B) triply-charged unmodified form of the same peptide sequence at 29 eV.

Figure 5. Tandem mass spectra of (A) doubly-charged sodiated peptide ion, VGYMHWYQQKPGK compared with (B) doubly-charged non-sodiated form of the same sequence.

Figure 6. Tandem mass spectra of triply-charged peptide ion with methionine alkylation, WQQGNVFSC(CAM)SVM(CAM)HEALHNHYTQK, at m/z 953.4344. Abundant peaks in the high m/z region arise from a neutral loss of C_3H_7NOS (105 Da) from parent and y ion series.

Figure 7. Selected ion chromatograms of oxidized methionine containing peptides illustrating the “bridge-like” peak connecting in-source and in-sample formed peptides. Unoxidized

methionine residues (Met255, Met361, Met431) are shown in a1, b1, and c1 while oxidized methionine residues (MetOX255, MetOX361, MetOX431) shown in a2, b2, and c2.

Figure 8. Comparison of relative abundance of each oxidized Met residue in the NISTmAb reported by this study and an interlaboratory study of identification and quantification of NISTmAb methionine oxidation.³⁶ NIST, this study; LAB1 – LAB3, Laboratory 1 to Laboratory 3 in Reference 36. LAB3 conducted a manual quantification of methionine oxidation. H, heavy chain. L, light chain.

Figure 9. Examples of rejected peptide spectra. (A) Rejected MS/MS spectrum identified as formyl₁SCD^{hex}KTHTCPPCPAPELLGGPSVFLFPPKPK at m/z 839.1618. (B) Rejected MS/MS spectrum identified as triply-charged tryptic peptide, D^{dehydrated}I^{cation:2Ca[III]}QMTQSPSTLSASVGDR at m/z 650.933.

Figure 10. Peptide identifications in a Humira® digest using the library developed in this work. Humira® was digested for 2 h after denaturing in 6 M guanidine at room temperature. Major, Classes 1 and 2 peptides; Glyco, N-linked glycopeptides; Oxid, Oxidized peptides; Deam, deamidated peptides; Glyca, Glycated peptides; Other, many analytical artifacts such as sodium adducts and formyl, and very long peptides; Semi, semi-tryptic peptides; Diag, miscleaved, and missed/over alkylated peptides.

Tables

Table 1. Twelve varieties of spectra in the NISTmAb library

Table 2. Number of distinct sequences, peptides, ions, and spectra for the heavy and light chains selected in the library that are derived from over six million peptide-spectrum matches (PSMs)

Table 3. Representative overalkylated peptides detected by 1D LC-MS/MS of the NISTmAb using Protocol 4 GuanRT/TCEP with 18 h of digestion

Table 4. Detailed analysis of 1D LC-MS peptide mapping data determined for the NISTmAb heavy and light chains. Median values in Column 4 - 8 were calculated across triplicate injections for each peptide under Protocol 4 GuanRT/TCEP with 18 h of digestion

Table 5. Library summary of a total of 24 modification categories detected in the NISTmAb tryptic peptides by combined results of all 13 sample preparation protocols

Table 6. Abundances and retention times of glycated peptides obtained from 1D and 2D LC-MS/MS analyses of 2 h NISTmAb digests. The 1D study includes 10 protocols excluding Protocols 3, 12, and 13 (see Materials and Methods), while 2D study uses only Protocol 2 GuanRT/DTT

Tables

Table 1. Twelve varieties of spectra in the NISTmAb library

	spectra varieties	# ions	# spectra
1	Unmodified major class peptide ^a	378	2239
2	N-linked glycosylated peptides ^b	247	1700
3	oxidized peptides	668	1945
4	deamidated peptides	245	996
5	N-/C terminal peptides	137	547
6	glycated peptides	93	656
7	formylated peptides ^c	117	483

8	adducted peptides	125	191
9	overalkylated peptides	30	105
10	other types of modified peptides ^d	206	642
11	diagnostic peptides ^e	466	2027
12	in-source or in-solution semi-tryptic peptides ^f	1007	3189

Note: ^a Variety 1 (e.g., no missed cleavages or expected missed cleavages). ^b Variety 2 contains glycopeptides from the NISTmAb glycopeptide library.¹⁷ ^c because formic acid was used in the protocol. ^d Variety 10 includes transpeptidation, dehydration, gln->pyro-glu, glu->pyro-glu, ammonia-loss, and others. ^e Variety 11 contains miscleaved and under/over alkylated peptides. ^e semi-tryptic peptides are produced in ESI (in-source) or in digestion (in-solution).

Table 2. Number of distinct sequences, peptides, ions, and spectra for the heavy and light chains selected in the library that are derived from over six million peptide-spectrum matches (PSMs)

protein	# sequences	# peptides	# ions	# spectra	# PSMs
heavy chain (total)	639	1,289	2,387	8,943	3,456,578
heavy chain (constant regions)	445	843	1,599	6,151	2,575,105
light chain (total)	317	593	973	3,665	2,962,003
light chain (constant region)	164	292	472	1,777	1,289,388

Note: peptide = sequence + modification, peptide ion = sequence + modification + charge state, spectra = sequence + modification + charge state + collision energy

Table 3. Representative overalkylated peptides detected by 1D LC-MS/MS of the NISTmAb using Protocol 4 GuanRT/TCEP with 18 h of digestion

peptide ^a	z	site	non-overalkylated		overalkylated	
			RT min	log(int)	RTdiff	% int
WQQGNVFSCSVMEALHNHYTQK	3, 4, 5	Met	47.88	11.95	-13.44	4.2
DIQMTQSPSTLSASVGD	2, 3	Met	48.04	11.32	-13.16	9.3
ESGPALVKPTQLTLTCTFSGFSLSTAGMSVGVIR	3	Met	77.42	10.96	-8.35	11.7
DTLMISR	1, 2	Met	34.01	11.50	-9.90	2.5
VGYMHWYQKPGK	2, 3, 4	Met	34.22	11.55	-11.56	1.9
VTNMDPADTATYYCAR	2, 3	Met	44.57	10.85	-12.91	6.2
VYACEVTHQGLSSPVTK	2, 3	N-term	35.80	11.76	1.03	0.3
FNWYVDGVEVHNAK	2, 3	N-term	48.81	11.82	1.45	0.2
STSGGTAALGCLVK	2	N-term	40.49	11.66	0.86	0.1
EPQVYTLPPSREEMTK	3	Met	38.72	11.22	-9.67	0.3
VDNALQSGNSQESVTEQDSK	2, 3	N-term	31.62	11.64	0.68	0.1
HKVYACEVTHQGLSSPVTK	3, 4	N-term	28.98	10.89	1.50	0.3
TPEVTCVVVDVSHEDPEVK	3	Ser	50.25	11.29	-0.49	0.1

Note: ^a all cysteines were alkylated. RT/min: retention time in min. log(int): log of summed peak area of all charge states. RTdiff: retention time difference (in min) of overalkylated and non-overalkylated

peptides, % int: percentage of total intensity of overalkylated and non-overalkylated peptides. Median RT, log(int), RTdiff, and % int were calculated from the triplicate analysis.

Table 4 Detailed analysis of 1D LC-MS peptide mapping data determined for the NISTmAb heavy and light chains. Median values in Column 4 - 8 were calculated across triplicate injections for each peptide under Protocol 4 GuanRT/TCEP with 18 h of digestion.

pep	position	m/z	error ppm	RT min	log(int)	relab	stdev	z ^f	Sequence ^{g,h}
Heavy Chain									
H1 ^a	1-5	599.351	0.7	33.4	11.4	0.2591	0.08	1	pQVTLR
H2	6-40	1234.296	0.4	77.4	10.9	0.0580	0.02	3	ESGPALVKPTQTLTLCTFSGFS LSTAGMSVGVWIR
H3 ^b	41-45	526.298	1.5	9.1	9.0	0.0012	0.00	1	pQPPGK
H4	46-59	596.970	0.3	73.5	11.3	0.1665	0.08	3	ALEWLADIWDDKK
H5	60-66	429.727	3.3	17.8	11.6	0.4052	0.14	2	HYNPSLK
H6	69-73	281.184	2.9	19.1	11.1	0.1674	0.00	2	LTISK
H7	74-77	450.220	1.5	9.0	8.2	0.0002	0.00	1	DTSK
H8	78-83	350.721	0.9	18.5	11.3	0.2795	0.00	2	NQVVLK
H9	84-99	924.898	0.9	44.6	10.8	0.0539	0.04	2	VTNMDPADTATYYCAR
H10	100-124	934.442	1.9	95.9	9.8	0.0047	0.00	3	DMIFNFYFDVWGQGTTVTVSSA STK
H11	125-136	593.827	0.8	44.6	11.8	0.5753	0.27	2	GPSVFPLAPSSK
H12	137-150	661.343	0.6	40.5	11.6	0.4027	0.22	2	STSGGTAALGCLVK
H13	151-213	1343.469	1.9	78.4	11.4	0.2597	0.01	5	DYFPEPVTVSWNSGALTSGVHT FPAVLQSSGLYSLSSVTVPSSSL GTQTYICNVNHKPSNTK
H14	214-217	259.158	1.2	8.5	8.5	0.0005	0.00	2	VDKR
H15	218-221	472.277	1.9	9.1	9.5	0.0031	0.01	1	VEPK
H16	222-225	509.202	0.1	9.0	7.9	0.0001	0.00	1	SCDK
H17	226-251	711.870	2.7	59.4	11.7	0.4976	0.06	4	THTCPPELGLGSPVFLFPPK PK
H18	252-258	418.221	0.6	34.0	11.4	0.2845	0.09	2	DTLMISR
H19	259-277	713.681	2.1	50.3	11.7	0.3922	0.14	3	TPEVTCVVVDVSHEDPEVK
H20	278-291	559.939	0.4	48.8	11.7	0.4987	0.10	3	FNWYVDGVEVHNAK
H21	292-295	501.314	1.8	9.0	7.9	0.0001	0.00	1	TKPR
H22 ^c	292-304	820.354	1.9	19.8	10.4	0.0210	0.01	4	TKPREEQYNSTYR
H23 ^c	296-304	1398.553	0.7	23.7	10.6	0.0390	0.02	2	EEQYNSTYR
H24	305-320	603.340	1.7	65.6	11.4	0.1978	0.01	3	VVSVLTVLHQDWLNGK
H25	321-323	439.219	1.3	9.2	9.6	0.0036	0.00	1	EYK
H26 ^d	324-325	307.144	3.1	5.4	8.3	0.0001	0.00	1	CK
H27	326-329	447.256	0.0	8.5	8.6	0.0006	0.00	1	VSNK
H28	330-337	419.755	0.3	29.7	11.6	0.3580	0.19	2	ALPAPIEK
H29	338-341	448.277	2.9	9.1	9.7	0.0671	0.02	1	TISK
H30	344-347	457.252	1.5	9.0	8.7	0.0004	0.00	1	GQPR

H31	348-358	643.841	2.5	38.4	11.6	0.3734	0.16	2	EPQVYTLPPSR
H32	359-363	319.147	2.2	11.0	10.0	0.0151	0.00	2	EEMTK
H33	364-373	581.319	1.0	43.9	11.6	0.3372	0.16	2	NQVSLTCLVK
H34	374-395	1272.570	1.4	63.6	10.9	0.0934	0.02	2	GFYPSDIAVEWESNGQPENNYK
H35	396-412	937.465	0.4	63.5	11.5	0.2794	0.02	2	TTPPVLDSDGSSFFLYSK
H36	413-417	288.174	1.7	16.0	11.0	0.1369	0.00	2	LTVDK
H37	420-442	701.072	0.8	47.9	11.6	0.3868	0.20	4	WQQGNVFSCSVMHEALHNHYT QK
H38 ^e	443-449	660.356	0.1	37.5	11.2	0.1849	0.06	1	SLSLSPG
H39	443-450	394.729	0.2	27.7	10.3	0.0185	0.01	2	SLSLSPGK
Light Chain									
L1	1-18	946.955	0.6	48.0	11.2	0.1472	0.10	2	DIQMTQSPSTLSASVGR
L2	19-28	541.269	1.2	23.8	11.1	0.1338	0.05	2	VTITCSASSR
L3	29-41	406.204	1.0	34.1	11.3	0.2879	0.00	4	VGVMHWYQQKPGK
L4	42-44	315.203	1.4	8.6	8.2	0.0002	0.00	1	APK
L5	45-52	476.771	0.0	38.2	11.4	0.2715	0.10	2	LLIYDTSK
L6	53-60	393.727	4.8	21.6	11.3	0.2557	0.14	2	LASGVPSR
L7	61-102	1121.506	1.2	84.5	10.1	0.0112	0.01	4	FSGSGSGTEFTLTISSLQPDFAT YYCFQGSQGYPFIFGGGTK
L8	103-106	488.308	1.1	13.1	9.9	0.0071	0.00	1	VEIK
L9	103-107	322.708	0.1	12.6	10.5	0.0424	0.00	2	VEIKR
L10	107-125	701.381	2.1	55.6	10.7	0.0527	0.02	3	RTVAAPSVFIFPPSDEQLK
L11	108-125	973.517	0.7	63.1	11.6	0.4367	0.10	2	TVAAPSVFIFPPSDEQLK
L12	126-141	899.451	1.0	70.8	11.4	0.2082	0.02	2	SGTASVVCLLNNFYPR
L13 ^d	142-144	347.193	5.0	5.6	9.3	0.0014	0.00	1	EAK
L14	145-148	280.663	1.4	19.3	11.1	0.1797	0.00	2	VQWK
L15	149-168	1068.489	0.6	31.6	11.3	0.1990	0.16	2	VDNALQSGNSQESVTEQDSK
L16	169-182	751.883	0.9	55.4	11.4	0.1865	0.11	2	DSTYLSLSTLTLTK
L17	183-187	625.283	2.7	11.1	10.3	0.0178	0.02	1	ADYEK
L18	188-206	536.026	0.0	28.8	10.8	0.0571	0.03	4	HKVYACEVTHQGLSSPVTK
L19	190-206	625.981	0.9	35.8	11.7	0.4485	0.19	3	VYACEVTHQGLSSPVTK
L20	207-210	523.262	1.3	11.7	10.4	0.0252	0.02	1	SFNR
L21	207-213	435.182	1.0	17.3	9.7	0.0057	0.00	2	SFNRGEC
L22	211-213	365.113	1.2	9.1	8.6	0.0004	0.00	1	GEC

Note: ^a pQ = protein N-terminal pyroglutamic acid. ^b pQ = peptide N-terminal pyroglutamic acid. ^c summed values of N-linked glycopeptides carrying asialylated, biantennary core-fucosylated glycans. ^d values obtained from 2 h of digestion. ^e the C-terminal Lys clip. ^f most abundant charge state. ^g all cysteine residues are alkylated. ^h four possible dipeptides were not detected, e.g., DR (heavy chain 67-68), AK (heavy chain 342-343), SR (heavy chain 418-419), HK (light chain 187-188), but present in peptides with miscleavages. **Pep**: tryptic peptide number. **m/z**: theoretical value. **Relab**: ratio of the ion abundance relative to the maximum ion of the run. **stdev**: standard deviations of relative abundances. **H22, L18, L21**: these peptides represented expected-miscleavages are included to complement the primary structure.

Table 5. Library summary of a total of 24 modification categories detected in the NISTmAb tryptic peptides by combined results of all 13 sample preparation protocols

modification	mass difference	site ^a	# ions	# spectra	origin
Oxidation	15.9949	M, W, K, H, P	668	2482	in-sample, in-source, in-column or in-digestion
Deamidated	0.9840	N, Q	245	996	in-sample or in-digestion
Formyl	27.9949	N-terminal, K, S, T	117	483	in-digestion
Hex	162.0528	N-terminal, K	93	656	in-sample or in-digestion
Gln->pyro-Glu	-17.0265	Q at N-terminal	70	291	in-sample or in-digestion
Cation:Fe[III]	52.9115	any	42	60	in-source
Glu->pyro-Glu	-18.0106	E at N-terminal	33	86	in-digestion
Carbamidomethyl	57.0215	M, N-terminal	30	105	in-digestion
Cation:Fe[II]	53.9193	any	29	43	in-source
Carbamyl	43.0058	N-terminal	28	66	in-digestion
Lys (transpeptidation)	128.0950	N-terminal	25	70	in-digestion
Cation:Na	21.9819	any	22	37	in-source
Lys loss (heavy chain)	-128.0950	C-terminal	17	72	in-sample
Arg (transpeptidation)	156.1011	N-terminal	17	49	in-digestion
Cation:Ca[II]	37.9469	any	17	29	in-source
Cation:2Na	43.9639	any	15	22	in-source
Carboxy	43.9898	M, N-terminal	13	23	in-digestion
Dehydrated	-18.0106	N-terminal	12	36	in-source or in-digestion
Pyro-carbamidomethyl	39.9949	C at N-terminal	5	17	in-digestion
Ammonia-loss	-17.0265	N	5	15	in-sample or in-digestion
Nitro	44.9851	Y	4	9	in-source
Acetyl	42.0106	H	4	5	in-sample or in-digestion
Dioxidation	31.9898	W	3	9	in-sample or in-source
Trioxidation	47.9847	W	1	1	in-sample or in-source

Note: ^a the N-terminal modification occurred during the digestion process refers to peptides, while the N-terminal modification detected in sample refers to protein. For example, Gln->pyro-Glu and Hex were found either in the protein N-terminal (in-sample) or at the peptide N-terminal (in-digestion), while other N-terminal modifications only occurred at the peptide level.

Table 6. Abundances and retention times of glycosylated peptides obtained from 1D and 2D LC-MS/MS analyses of 2 h NISTmAb digests. The 1D study includes 10 protocols excluding Protocols 3, 12, and 13 (see Method), while 2D study uses only Protocol 2 GuanRT/DTT.

region	site	peptide sequence	2D LC-MS	1D LC-MS
--------	------	------------------	----------	----------

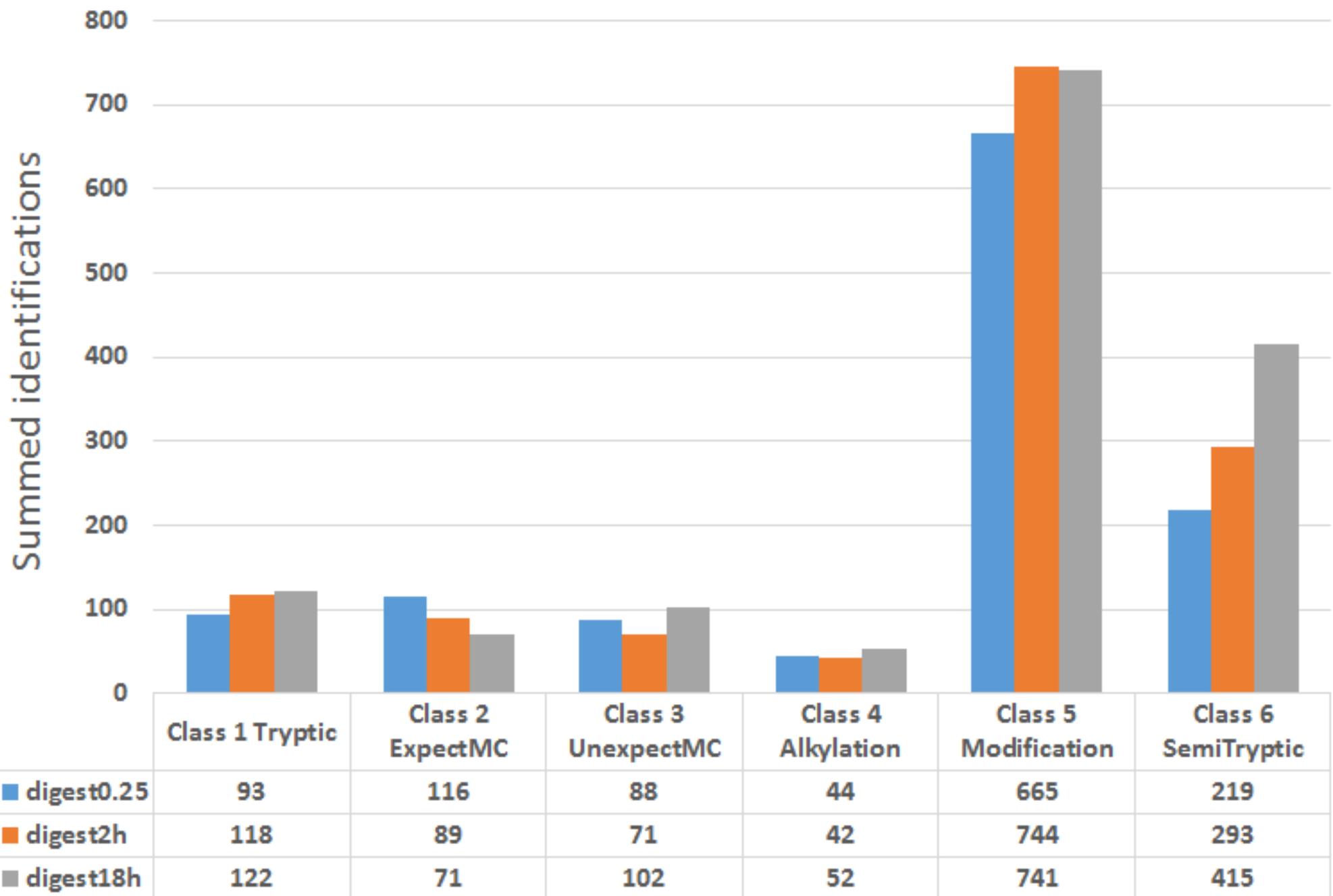
			z^c	log(int)	RT min	RT diff	# runs	stdev	log(int)	RT min	RT diff	# runs	stdev
heavy chain: V and CDR2 (58, 66)	58	ALEWLADIWWD DKK	2,3	8.85	81.66	-0.46	13	0.20	7.32	70.96	-0.75	7	0.03
	66	HYNPSLKDR	2,3	9.29	26.33	0.17	14	0.08					
	73	LTISK ^a	2	8.21	30.14	2.48	13	0.33					
		LTISKDTSKNQV VLK	2,3,4	9.75	41.28	-0.06	18	0.02	7.90	28.33	-0.18	12	0.04
		LTISKDTSK	2,3	9.68	26.44	0.24	13	0.13	7.89	15.05	0.37	9	0.10
	77	DTSKNQVVLK	2,3	8.85	30.55	0.61	17	0.06	7.56	17.43	0.24	9	0.03
83	NQVVVKVTNMD PADTATYYCAR	2,3,4	7.99	62.11	-0.16	8	0.05						
heavy chain: CH ₁	136	GPSVFPLAPSSK ^a	2	8.01	54.17	-0.43	17	0.08					
		GPSVFPLAPSSKS TSGGTAALGCLV K	2,3	8.82	65.90	-0.48	18	0.02	8.15	53.78	-0.47	11	0.04
hinge	225	SCDKTHTCPPCP APELLGGPSVFLF PPKPK	3,4,5	9.57	64.02	-1.12	17	0.23	8.81	54.69	-0.35	21	0.09
heavy chain: CH ₂	251	PKDTLMISR	2,3	7.91	38.46	-0.03	18	0.03	7.32	26.15	-0.01	3	0.05
	291	FNWYVDGVEVH NAKTKPR	3,4,5	9.05	50.59	-0.18	17	0.05					
	320	VVSVLTVLHQD WLNKEYK	2,3,4	8.59	70.02	-0.41	18	0.12	8.19	59.47	-0.31	25	0.05
	323	VVSVLTVLHQD WLNKEYKCK	3,4,5	9.13	66.65	-0.18	18	0.02	8.23	54.18	-0.20	3	0.02
	325	CKVSNKALPAPI EK	3,4	8.84	37.01	0.01	18	0.03	7.79	23.33	0.04	10	0.04
	329	VSNKALPAPIEK	2,3	10.51	38.63	-0.02	18	0.01	8.76	24.94	-0.04	25	0.02
	337	ALPAPIEK ^a	2	8.79	40.37	0.98	18	0.25					
		ALPAPIEKTISK	2,3	9.17	46.92	-0.13	17	0.05	7.92	33.67	-0.10	5	0.07
heavy chain: CH ₃	363	EEMTKNQVSLTC LVK	2,3	8.42	54.25	-0.19	16	0.03					
	395	GFYPSDIAVEWE SNGQPENNYKTT PPVLSDSDGSFFLY SK	3,4	7.82	84.09	-0.32	3	0.01					
	417	LTVDKSR	2,3	8.36	24.62	0.70	15	0.16					
	442	WQQGNVFSCSV MHEALHNHYTQ KSLSPG	3,4,5	8.92	65.55	-0.19	17	0.08					
	450	SLSLSPGK	2	8.41	38.96	0.17	17	0.03					
light chain: N- terminus	1	DIQMTQSPSTLSA SVGDR	2,3	8.71	59.08	0.73	18	0.12					
light chain: V and CDR2 (52)	19	VTITCSASSR ^b	2	7.75	36.21	0.52	18	0.09					
	38/4 1	VGYMHWYQQKP GK	2,3,4	8.75	45.13	-0.30	18	0.10					
		VGYMHWYQQKP GKAPK	3,4,5	9.41	42.02	-0.21	18	0.06	8.31	28.92	-0.14	14	0.07
	44	APKLLIYDTSK	3	7.81	48.11	-0.21	17	0.04					

	52	LLIYDTSK ^a	2	7.93	47.95	-0.49	18	0.10					
		LLIYDTSKLASG VPSR	2,3	9.13	59.98	-0.19	18	0.03	8.40	47.96	-0.21	19	0.03
	53	LASGVPSR ^b	2	7.81	33.79	1.65	18	0.20					
	102	YCFQSGYPFTF GGGTKVEIKR	3,4	8.30	63.28	-0.18	16	0.05					
	106	VEIKR	1,2	7.96	23.02	0.56	8	0.24					
light chain: C	125	TVAAPSVFIFPPS DEQLKSGTASVV CLLNNFYPR	3,4	9.42	94.81	-0.10	17	0.10					
	126	SGTASVVCLLNN FYPR ^b	2,3	8.14	79.24	-0.98	2	0.01					
	144	EAKVQWK	2,3	8.88	30.05	0.69	16	0.13					
	148	VQWKVDNALQS GNSQESVTEQDS K	2,3,4	9.53	53.52	-0.01	18	0.04	7.64	40.98	-0.03	18	0.03
		VQWKVDNALQS GNSQESVTEQDS KDSTYLSSTLTL SK	4	8.75	66.83	-0.09	18	0.04					
	168	VDNALQSGNSQE SVTEQDSKDSTY SLSSTLTLTK	3,4	8.97	63.28	-0.50	17	0.14					
		VDNALQSGNSQE SVTEQDSKDSTY SLSSTLTLTKADY EK	3,4	9.98	62.44	-0.31	18	0.02	8.51	50.22	-0.39	15	0.05
		VDNALQSGNSQE SVTEQDSKDSTY SLSSTLTLTKADY EKHK	4,5,6	9.27	58.00	-0.23	18	0.03	8.17	45.14	-0.17	9	0.06
	182	DSTYLSSTLTLK KADYEK	3	8.64	62.27	-0.38	15	0.05					
	187	ADYEKHKVYAC EVTHQGLSSPVT K	2,3,4, 5,6	10.48	43.95	-0.02	18	0.01	8.83	30.62	-0.03	18	0.03
	189	HKVYACEVTHQ GLSSPVTK	3,4	9.14	39.30	-0.13	9	0.09	8.10	26.75	-0.10	19	0.04
	206	VYACEVTHQGLS SPVTK ^a	3	8.05	45.93	-0.54	18	0.03					
		VYACEVTHQGLS SPVTKSFNRGEC	3,4	8.83	52.21	-0.08	18	0.03	7.84	39.29	-0.07	3	0.03

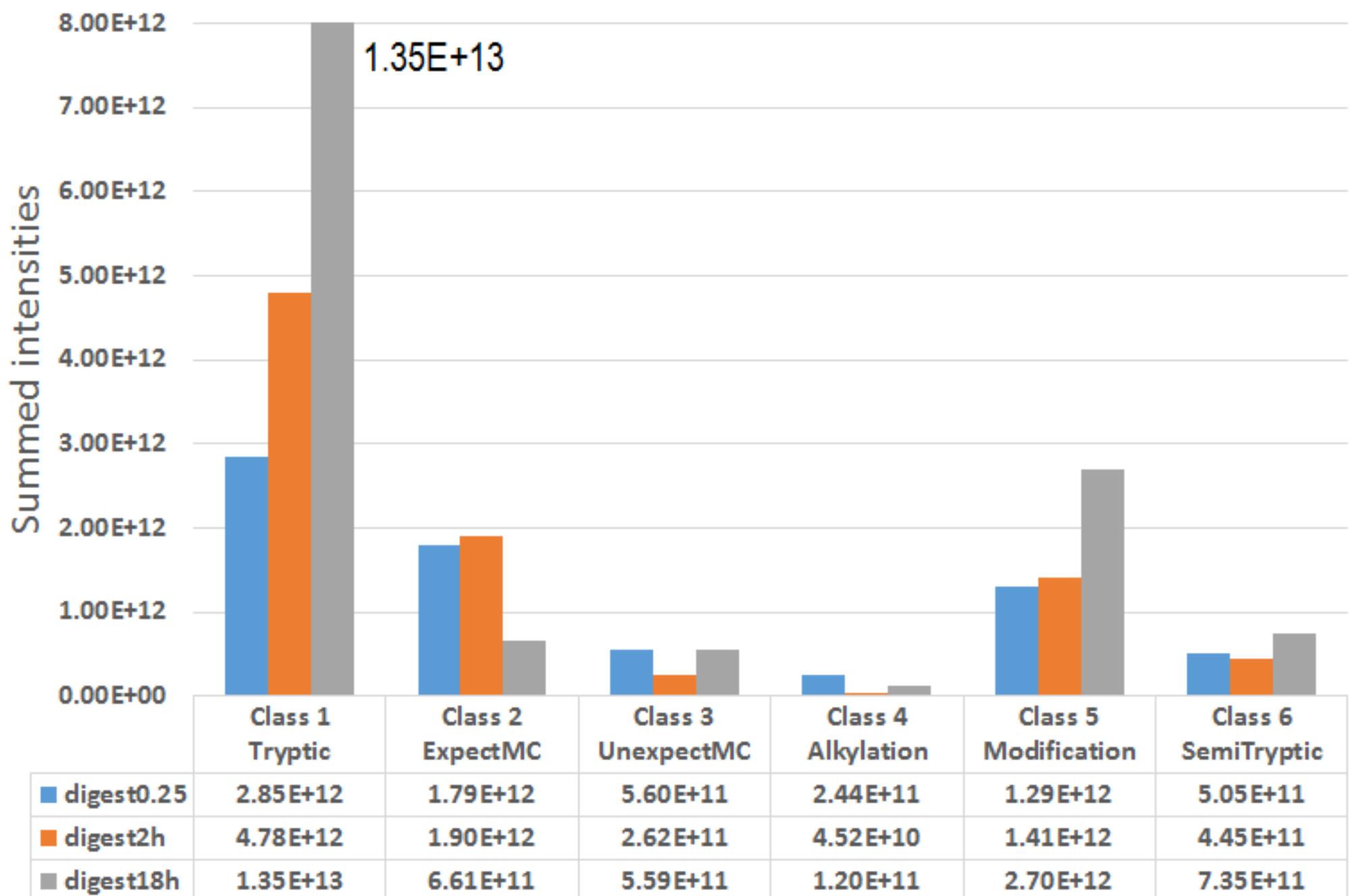
Note: ^a Glycation was detected on the peptide C-terminal lysine residue. ^b Glycation was detected on the peptide N-terminus. Both cases (a and b) were suspected to occur during or after digestion. ^c All charge states detected for this peptide with the most abundant charge shown in boldface. All cysteines were alkylated. region: **V** - variable domains, **C** – constant domains, **CH₁**, **CH₂**, **CH₃** - constant domains 1, 2 and 3, **CDR2** - complementarity-determining region 2; log(int): median absolute ion intensity (total peak

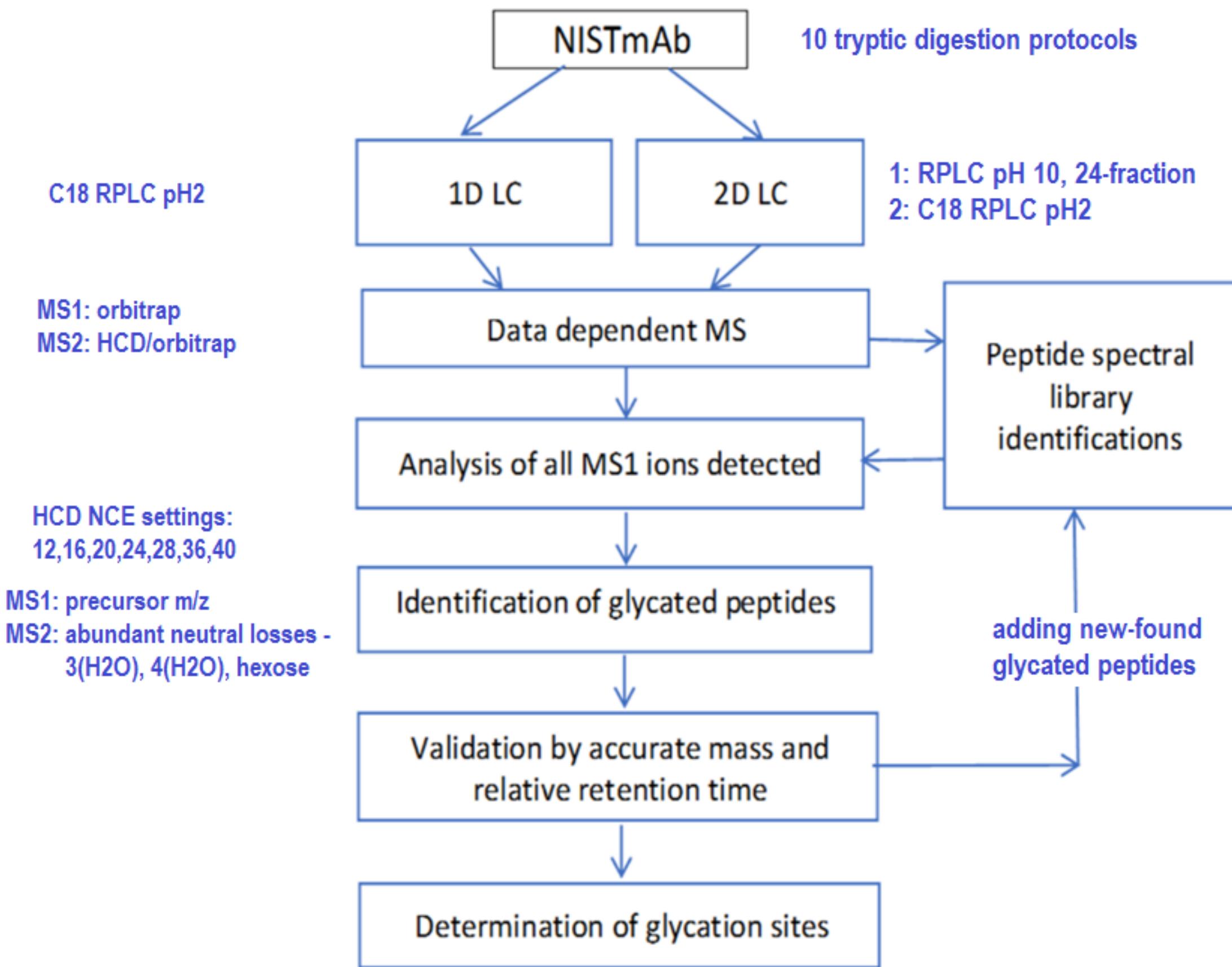
area) of the glycosylated peptide based on summing up all charge states; RT: median retention time (min);
RTdiff: glycosylated RT – unglycosylated RT; stdev: the standard deviation of RTdiff (min);

A summed peptide identifications of six classes



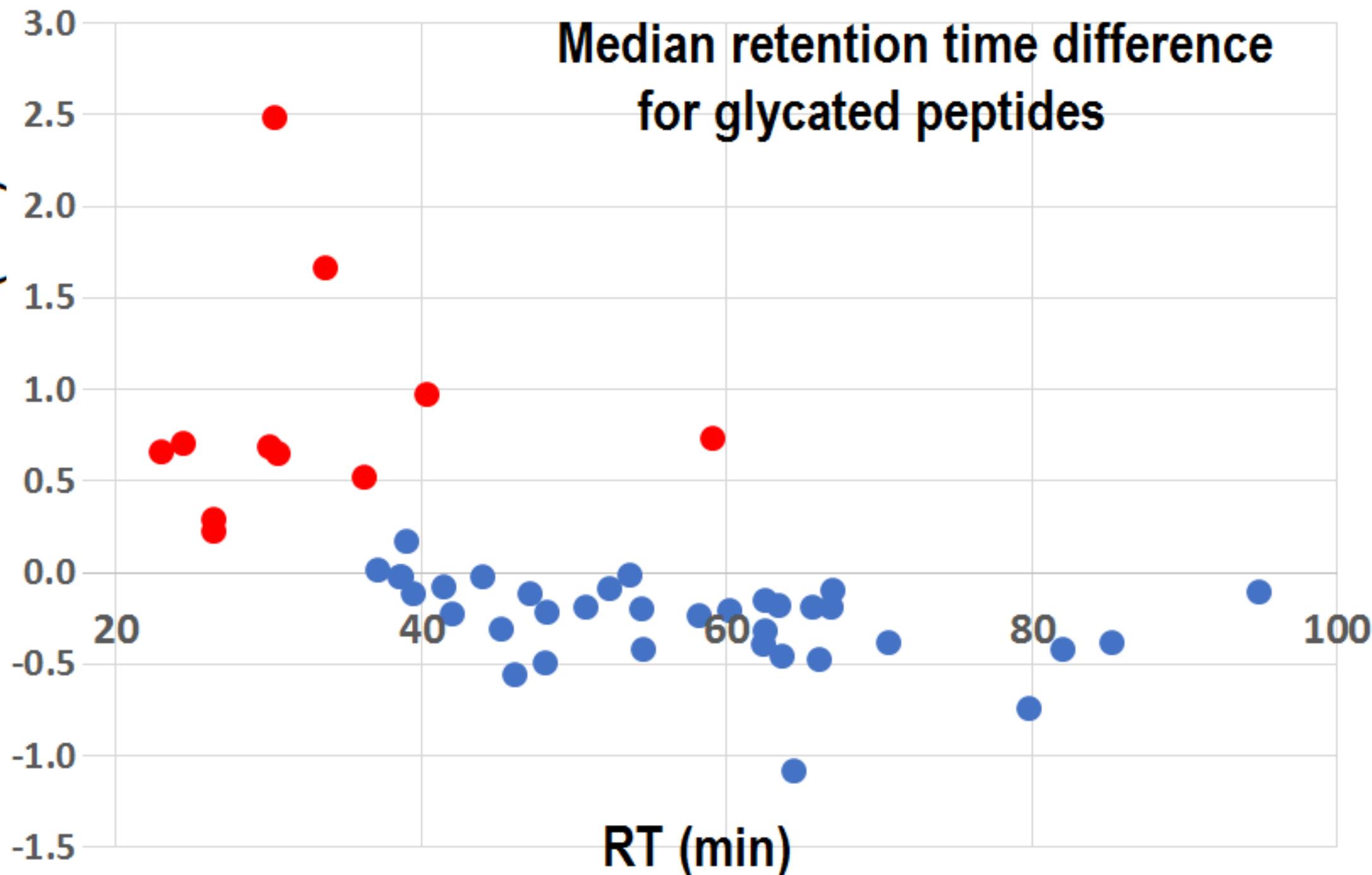
B summed peptide abundances of six classes



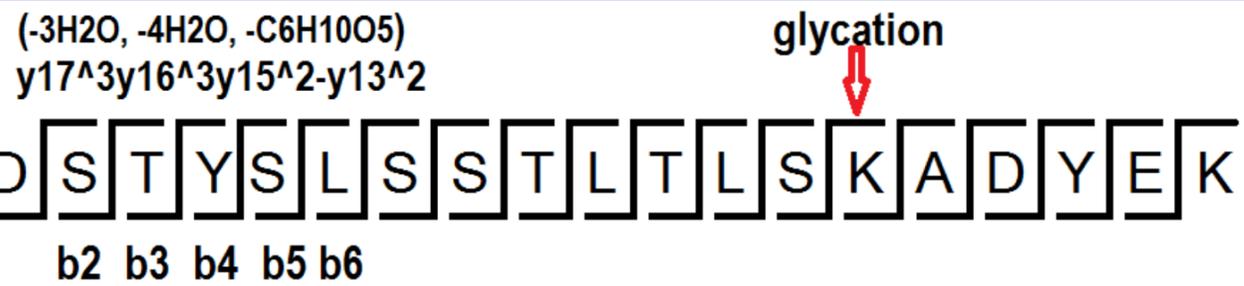


Median retention time difference for glycosylated peptides

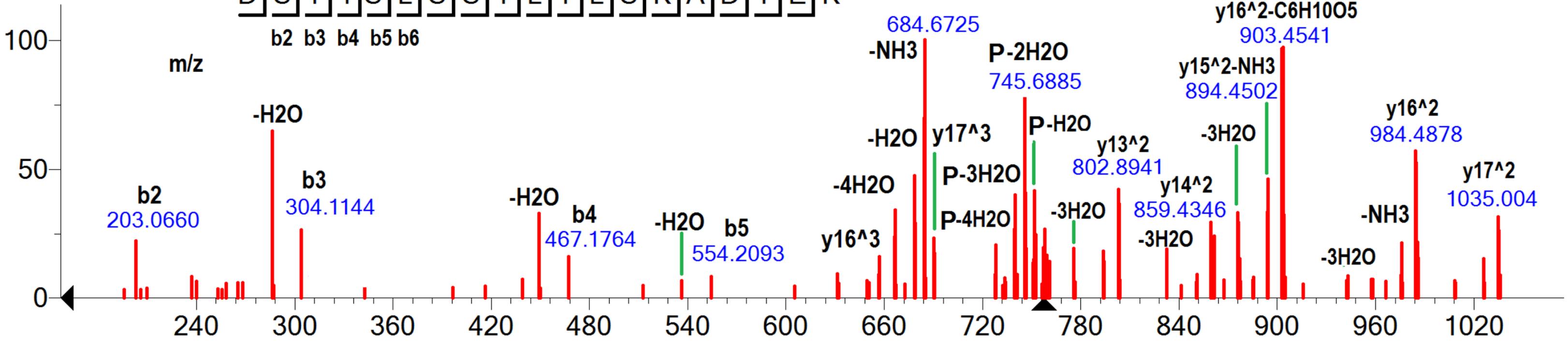
RT Difference (min)



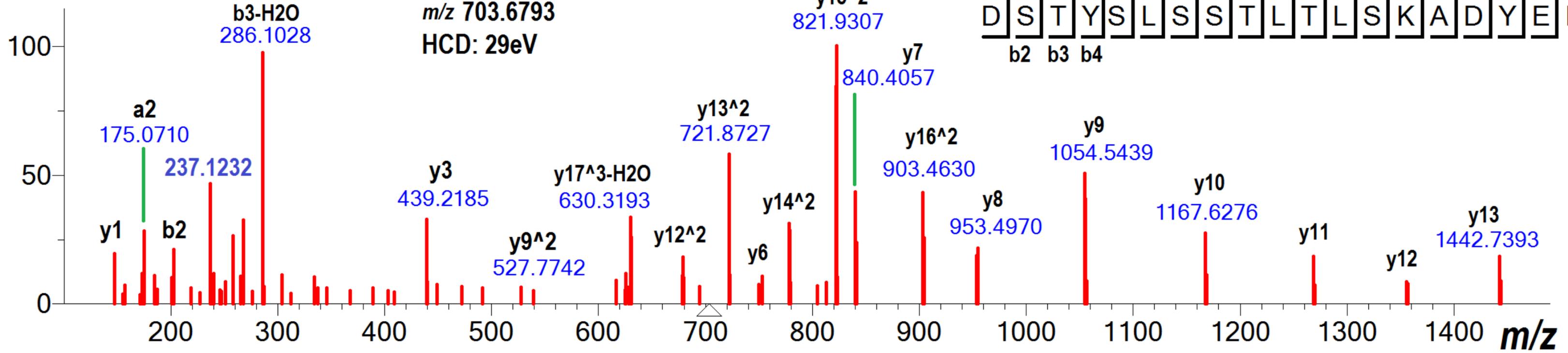
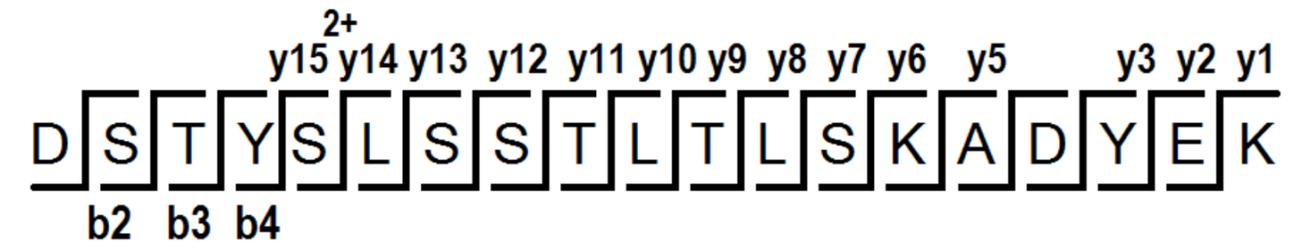
A. 3+ glycated ion
m/z 757.6969
 HCD: 21eV

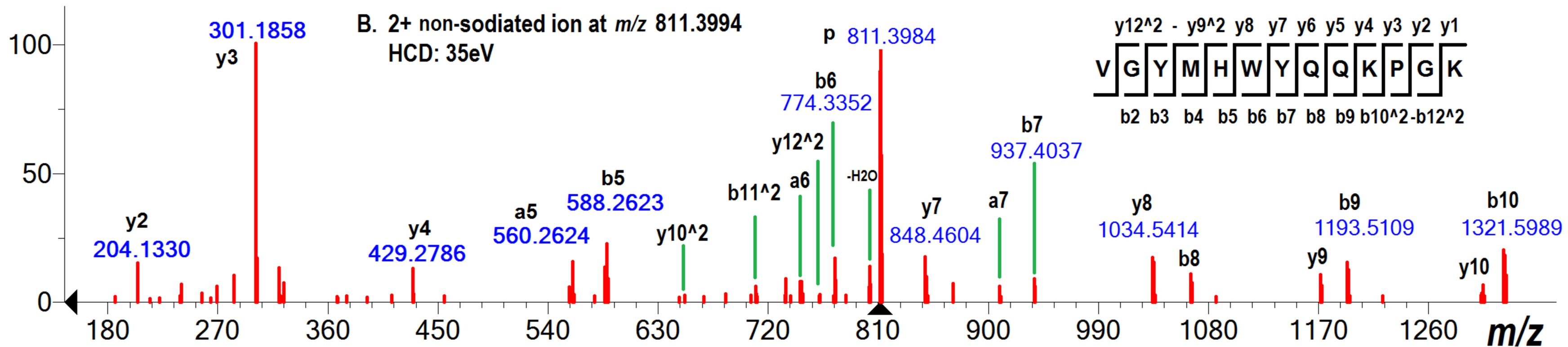
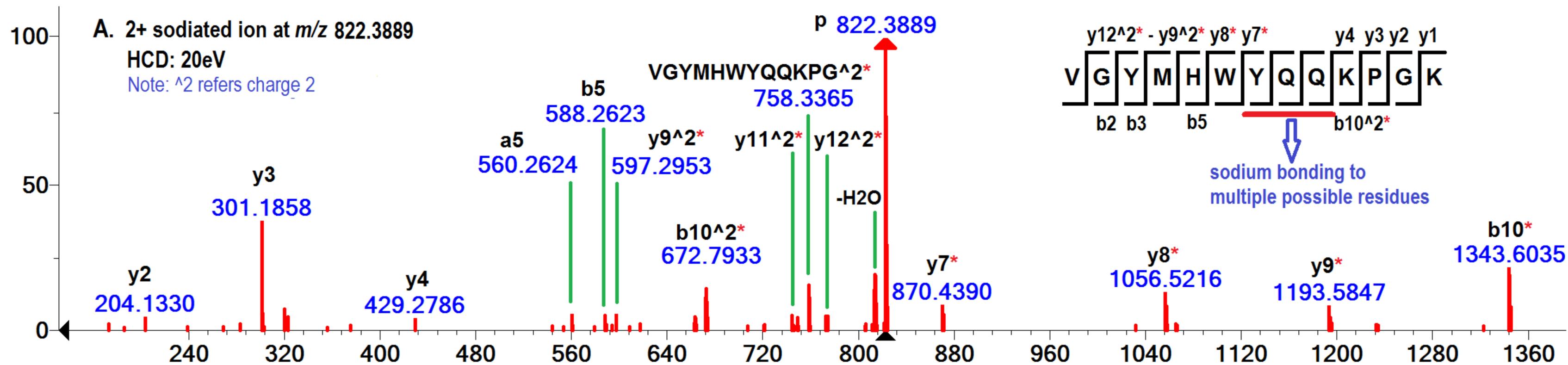


Note: ^2 refers charge 2



B. 3+ unglycated ion
m/z 703.6793
 HCD: 29eV





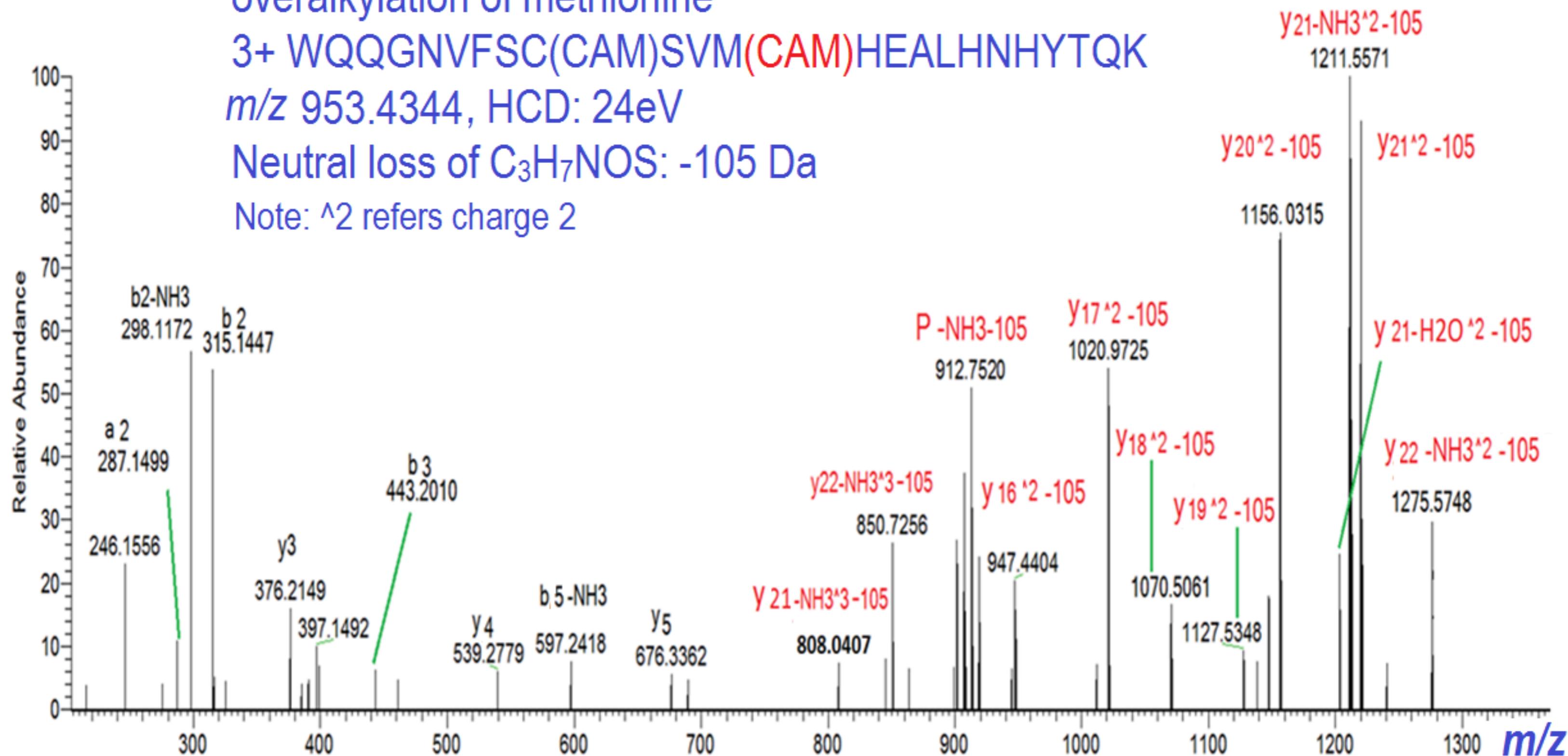
overalkylation of methionine

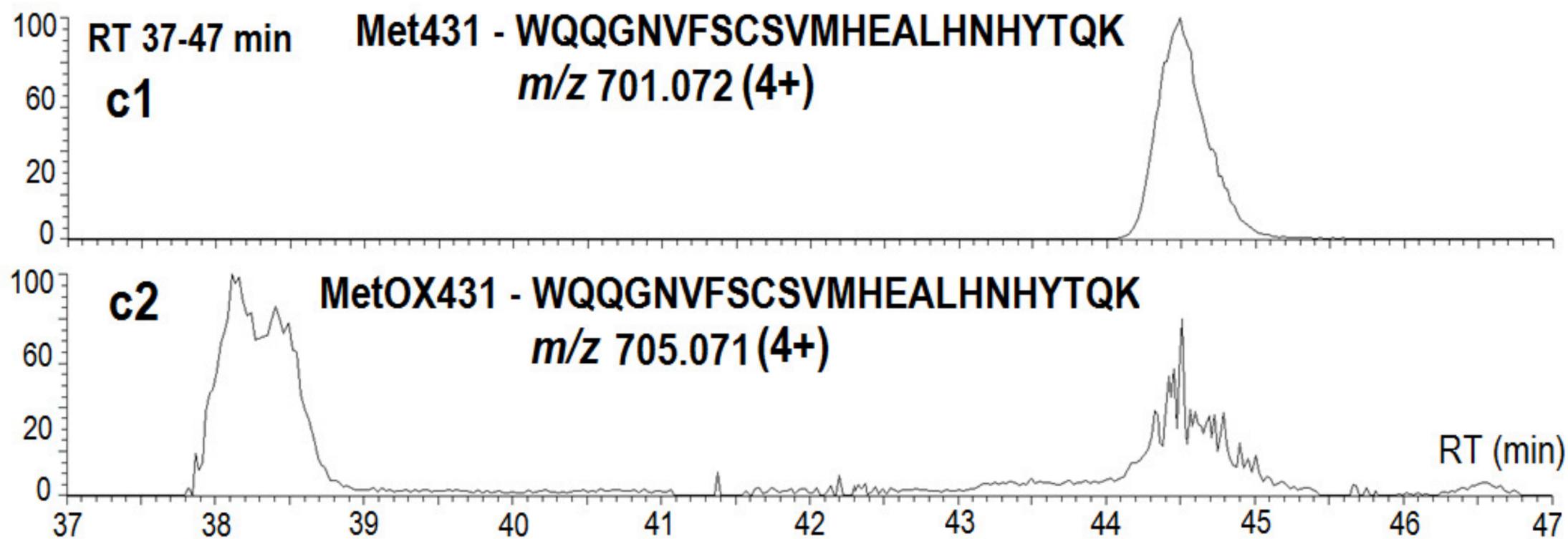
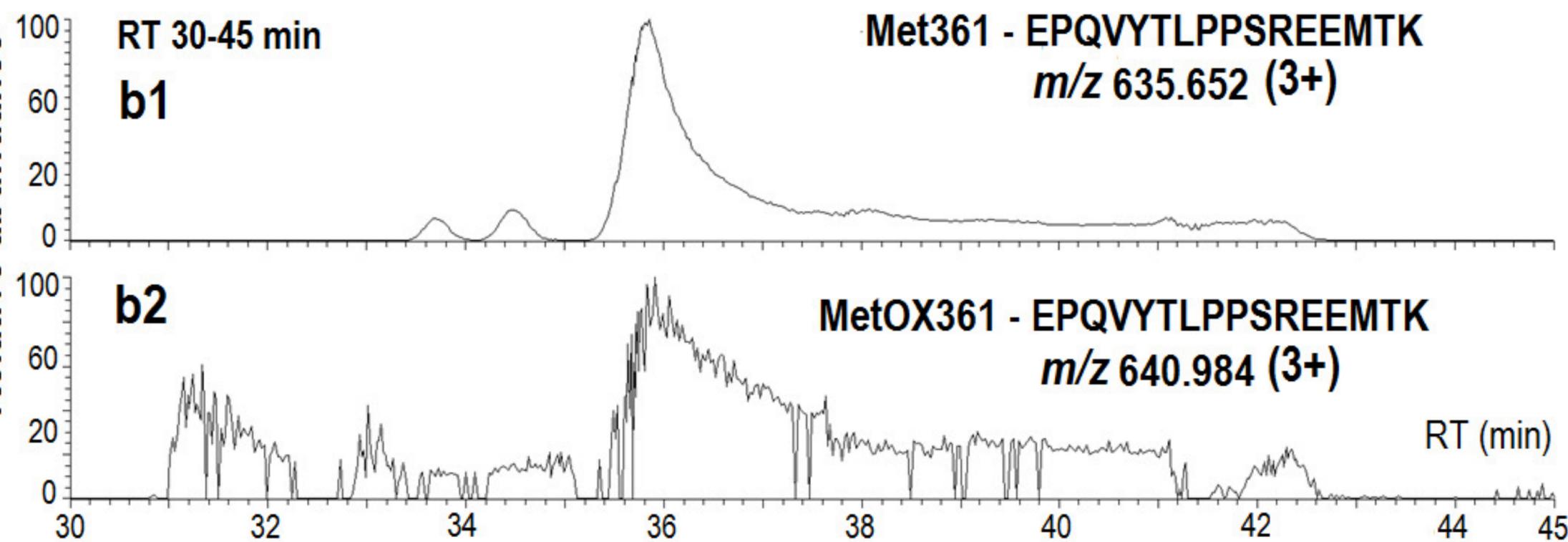
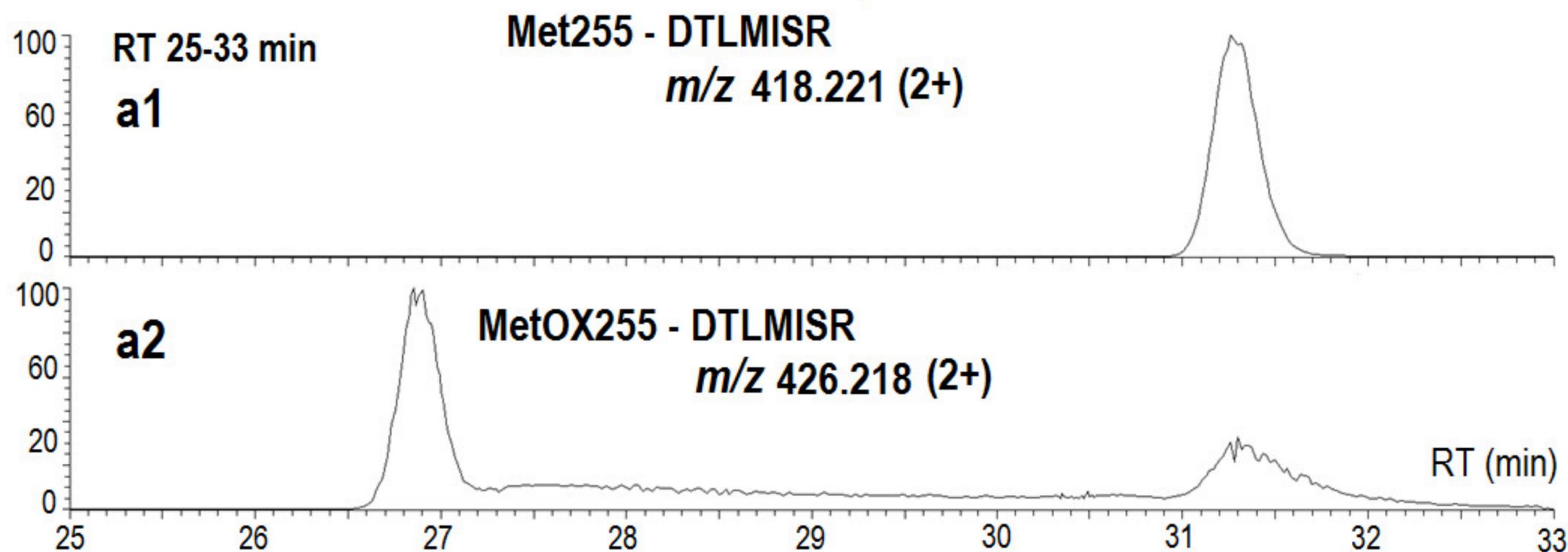
3+ WQQGNVFSC(CAM)SVM(CAM)HEALHNHYTQK

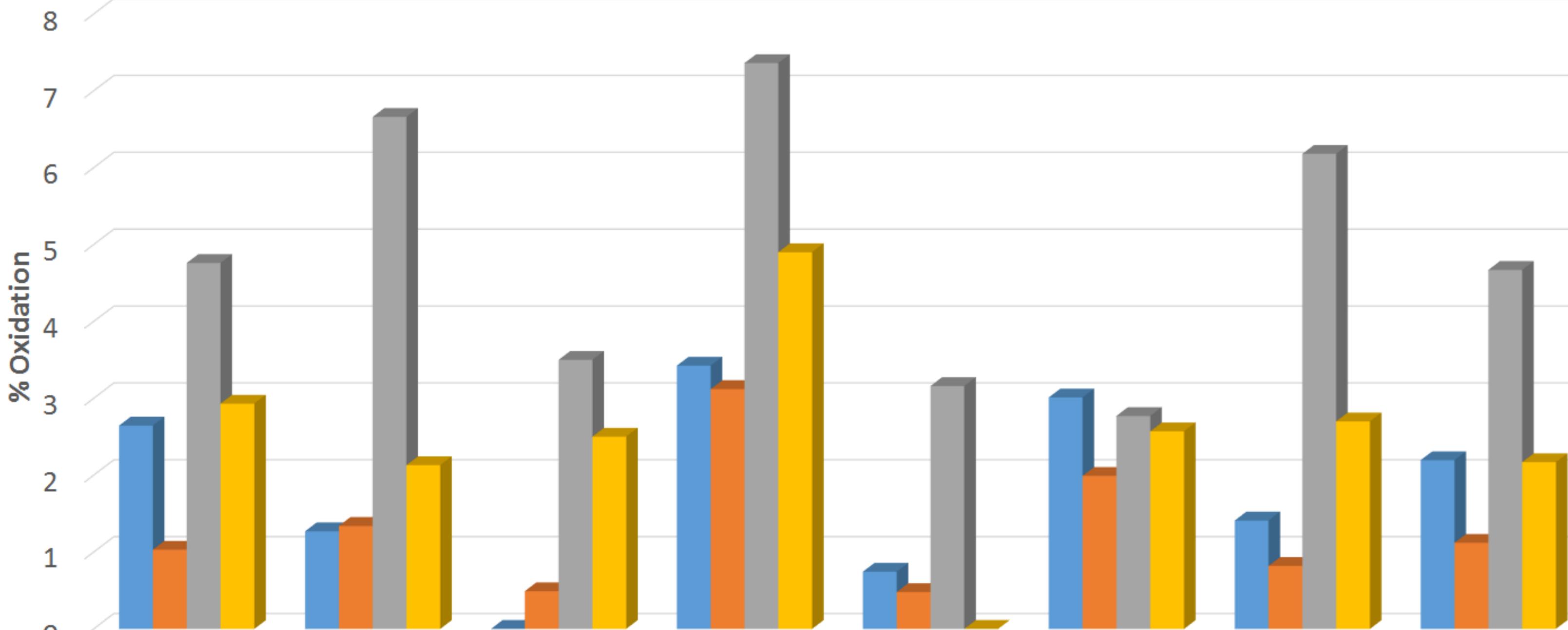
m/z 953.4344, HCD: 24eV

Neutral loss of C_3H_7NOS : -105 Da

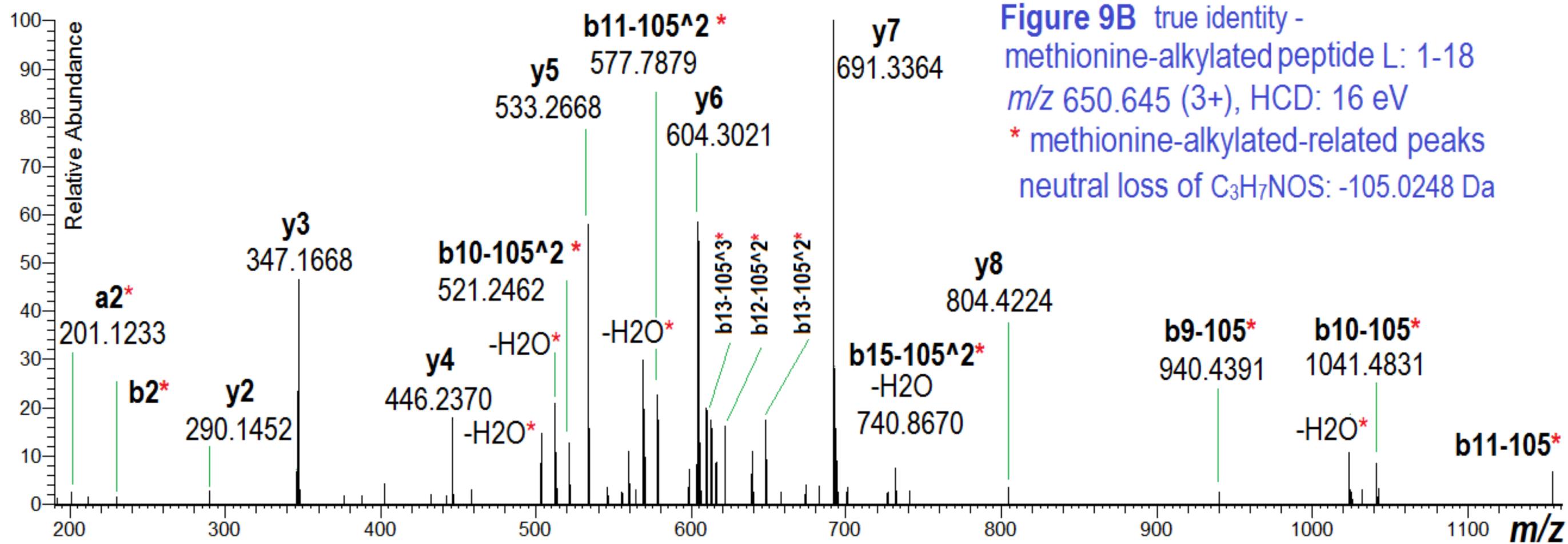
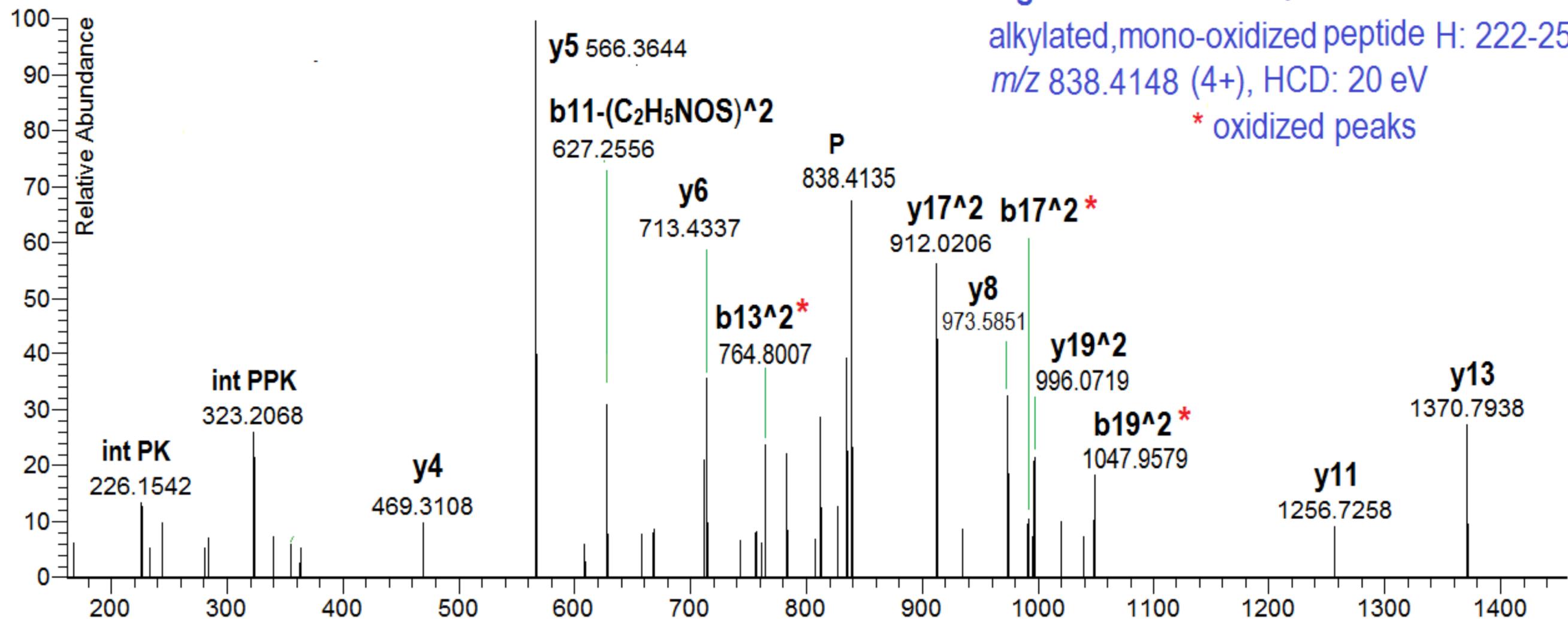
Note: ^2 refers charge 2



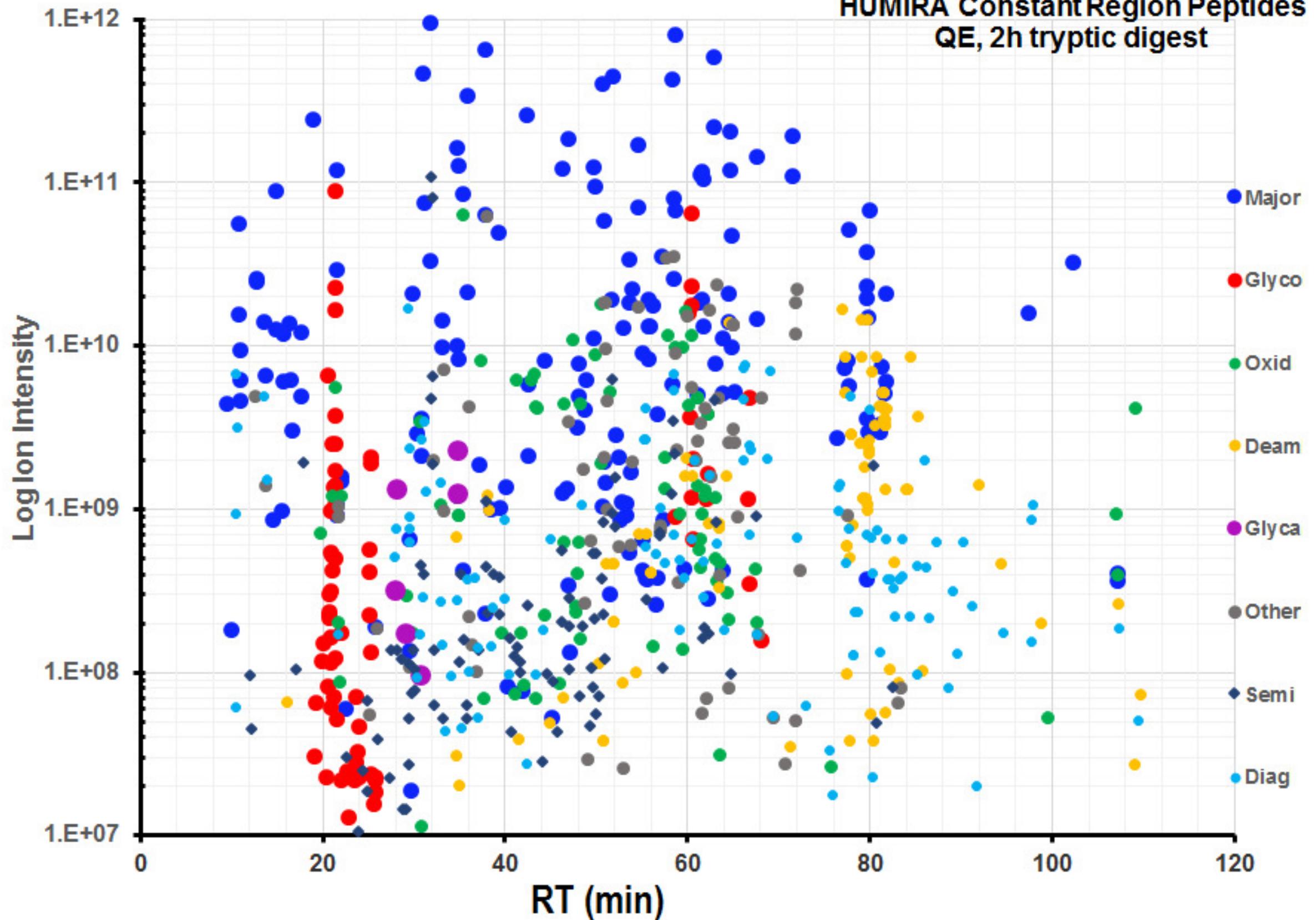




	H:Met34	H:Met87	H:Met101	H:Met255	H:Met361	H:Met431	L:Met4	L:Met32
■ NIST	2.65	1.27	0.00	3.43	0.74	3.01	1.41	2.20
■ LAB1	1.03	1.34	0.49	3.12	0.48	1.99	0.82	1.12
■ LAB2	4.76	6.66	3.50	7.36	3.16	2.77	6.18	4.67
■ LAB3	2.93	2.13	2.50	4.90	0.00	2.57	2.70	2.17



HUMIRA Constant Region Peptides
QE, 2h tryptic digest



The NISTmAb Tryptic Peptide Spectral Library for Monoclonal Antibody Characterization

Qian Dong*, Yuxue Liang, Xinjian Yan, Sanford P. Markey, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Tallat H. Bukhari, and Stephen E. Stein

Biomolecular Measurement Division, National Institute of Standards and Technology,
100 Bureau Drive, Stop 8362, Gaithersburg, Maryland 20899, United States

Supplementary Material Document S-1. NISTmAb Sequence Summary

Variable Fab—Constant Fab—Hinge—Constant Fc—secretory tail

>sp|NHVY|NH_NISTmAb heavy chain

QVTLRESGPALVKPTQTLLTCTFSGFSLSTAGMSVGVIRQPPGKALEWLADIWDDKKHYNPSLKDRLTISKDTSKNQVVLK
VTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGAL
TSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVKPKSCDKTHTCPPCPAPELLGGPSVFLFPPK
PKDTLMI SRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP
APIEKTI SKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTV
DKSRWQQGNV FSCSV MHEALHNHYTQKSLSLSPGK

>sp|NLGT|NL_NISTmAb light chain

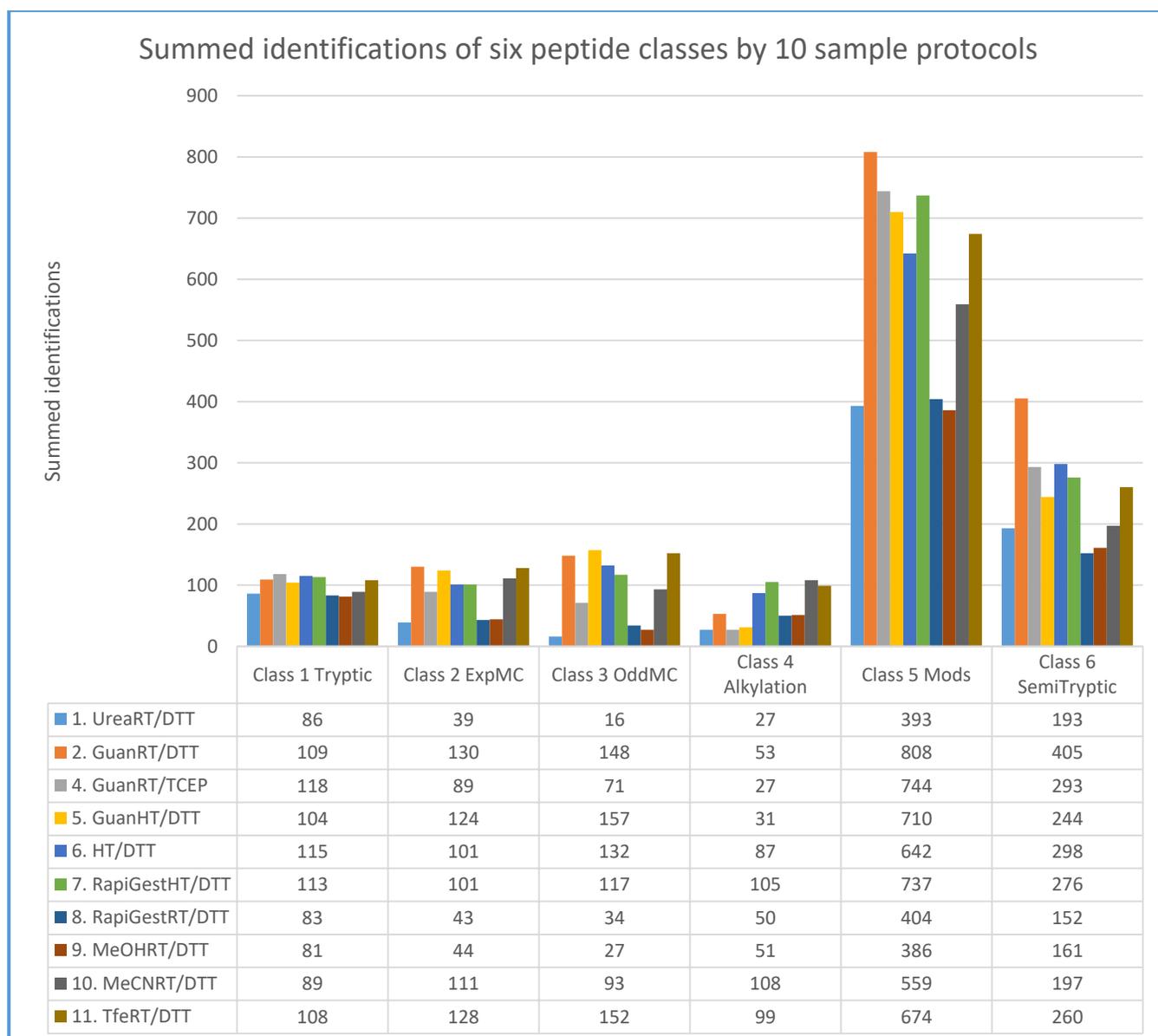
DIQMTQSPSTLSASVGVDRVITITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISLQPDFA
TYYCFQGSGLYPFTFGGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQD
SKDSTYLSLSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Supplementary Material Table S-1. Parameters for 13 protocol variations used during denaturing, reduction, and alkylation of the NISTmAb. Each sample was digested with trypsin for 0.25 h, 2 h and 18 h. 10-sample protocols (not including Protocols 3, 12, and 13) for the NISTmAb analysis were compared using six peptide classes (**Supplementary Material Figure S-1**).

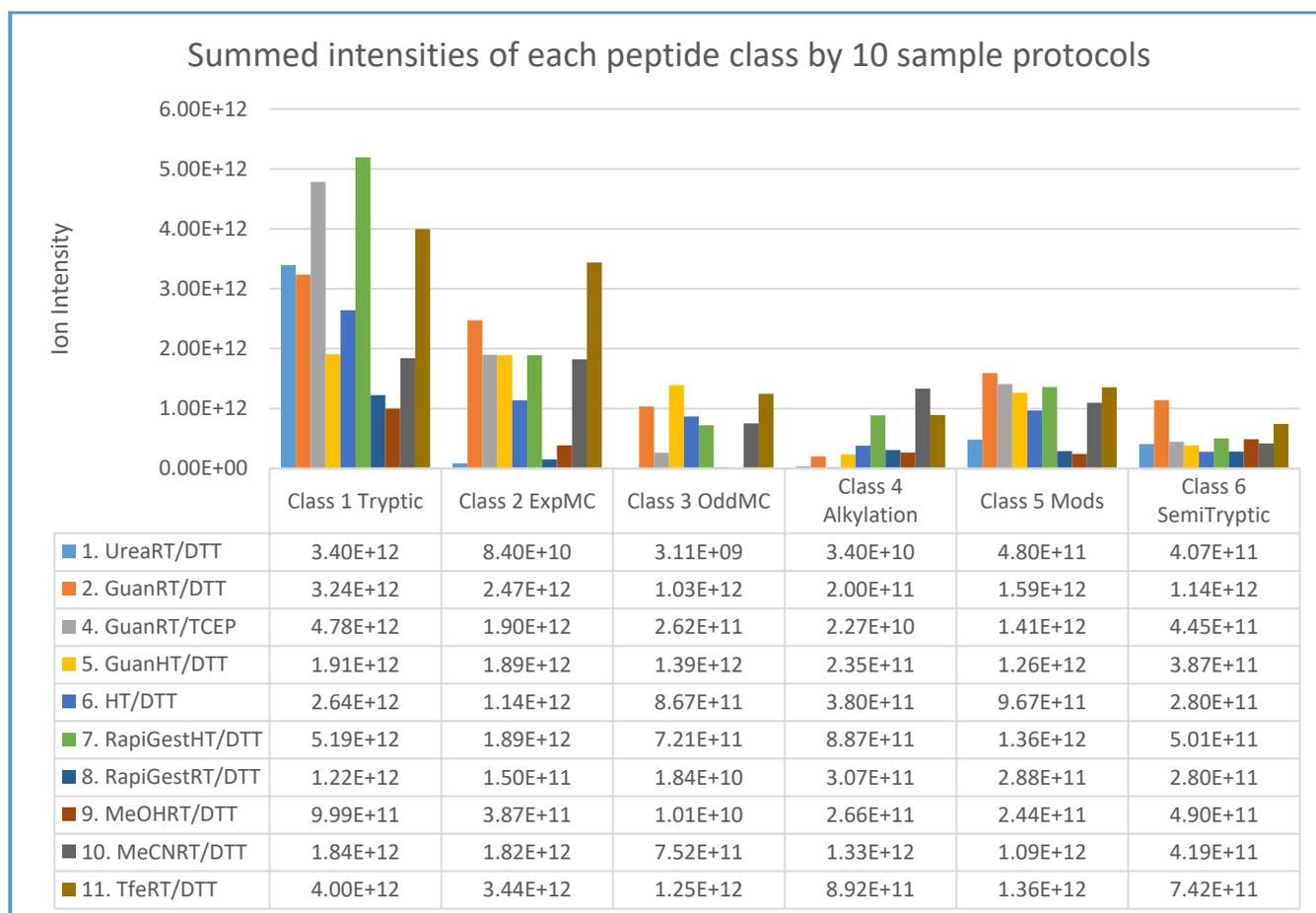
digestion	denaturation (1h)	temperature	reduction	alkylation 1h, RmTemp	IAA quenched (1h)	digestion (37°C)	
1. <i>UreaRT/DTT</i>	6M Urea	RmTemp	DTT	IAA	DTT	10 ug of sequencing-grade trypsin	
2. <i>GuanRT/DTT</i>	6M Guanidine	RmTemp	DTT		DTT		
3. <i>GuanRT/No2DTT</i>					not quenched		
4. <i>GuanRT/TCEP</i>			TCEP		not quenched		
5. <i>GuanHT/DTT</i>			85 °C		DTT		DTT
6. <i>HT/DTT</i>			heating (HT)		85 °C		DTT
7. <i>RapiGestHT/DTT</i>	0.2% RapiGest	85 °C					
8. <i>RapiGestRT/DTT</i>		RmTemp					
9. <i>MeOHRT/DTT</i>	20% methanol	RmTemp					
10. <i>MeCNRT/DTT</i>	20% acetonitrile						
11. <i>TfeRT/DTT</i>	50% TFE						
12. <i>GuanRT/DTT/CheapTrypsin10</i>	6M Guanidine	RmTemp	DTT		DTT	10 ug of Sigma T1426 trypsin	
13. <i>GuanRT/DTT/CheapTrypsin50</i>	6M Guanidine					50 ug of Sigma T1426 trypsin	

Supplementary Material Figure S-1. Comparison of ten tryptic digestion protocols using peptide class analysis. The protocol number and name can be found in **Supplementary Material Table S-1.**

A. Summed number of peptide identifications of each class obtained from 10 protocols at 2 h of digestion



B. Summed intensities of six peptide classes obtained by 10 sample protocols at 2 h of digestion



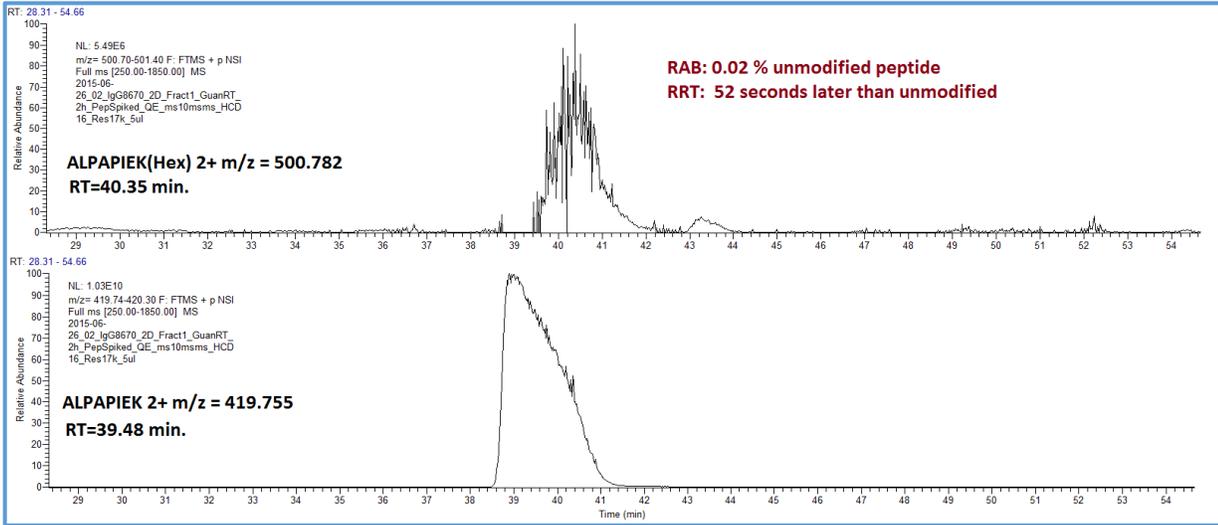
Note: Comparison of ten tryptic digestion protocols using peptide class analysis.

The analysis of peptide classes uses both the number and ion intensities of confident identifications of all peptide classes representing different digestion events. To demonstrate the utility of this method in selecting the preferable digestion conditions under which maximum abundances can be achieved for detailed peptide and PTM analyses of NISTmAb, Figure S-1 shows the results of comparison of 10 of 13 digestion protocols (see Method and Supplementary Material Table S-1) that have been widely used in the fields of proteomics and biopharmaceuticals. The analysis clearly revealed that both the summed number and abundances for each peptide class produced by different methods vary widely, except the number of identifications in Class 1, which suggested that changes in digestion protocols indeed greatly affect peptide identifications and abundances, especially for low abundance peptides and modifications.

We compared the digestion profiles of these 10 protocols and found that two protocols using GuanRT/TCEP and RapiGestHT/DTT (Method 4 and 7) produced a higher number as well as maximum abundance of major tryptic peptides for Class 1 and suitable number and abundance of tryptic peptides with expected missed-cleavage for Class 2. Both methods performed reasonably well for diagnostic classes (Class 3, 4, 5, and 6) where the percent abundance of these peptides were observed in the lower range (e.g. 3% to 8 %), although Method 7 showed higher level of incomplete digestion and alkylation relative to that of Method 4. Methods 1, 4, 5, 6, 8, and 10 consistently produced low yields in overall products of tryptic digestion, resulting in low number and abundance of detected peptides in all classes. Methods 2 and 11 were associated with lower digestion efficiency and contaminant as shown in high percentage of missed cleavage, semitryptic peptides, and artificial modifications as well. Another observation worth noting was that over half of the methods (Methods 6-11) caused unusually high percentage of incomplete alkylation on Cys from 6% to 18% as compared to less than 3% of all Guanidine-containing methods.

Supplementary Material Figure S-2. MS1 and MS2 spectra for an example of the new-found glycosylated peptide, ALPAPIEK, containing glycosylation of the C-terminal lysine. As shown in the Table 6, unlike common lysine-glycosylation that produce missed cleavages, the new-found glycosylated peptides contain no miscleaved Lys residues and their unmodified native peptides are major tryptic peptide products (Class 1). These glycosylated peptides were seen in almost all 2D runs and exhibit similar relative RT shifts described above. However, their abundance relative to unmodified forms ranged from 0.003% to 0.04%, suggesting that only very low-level glycosylation reactions occur at N-terminal amino acids and the C-terminal lysine of peptides. The origin of these reactions is unknown, may occur during protein digestion because the glycosylation of these sites cannot occur during the biomanufacturing process.

A. Selected ion chromatograms of identified glycosylated and unglycosylated peptides from the same 2D run.



B. Annotated MS2 spectrum of glycosylated peptide, ALPAPIEK.

