Published in final edited form as:

J Am Soc Mass Spectrom. 2013 July ; 24(7): 1072–1081. doi:10.1007/s13361-013-0624-y.

Unusual fragmentation pathways in collagen glycopeptides

Irina Perdivara^{1,*}, Lalith Perera², Marnisa Sricholpech³, Masahiko Terajima⁴, Nancy Pleshko⁵, Mitsuo Yamauchi⁴, and Kenneth B. Tomer^{1,*}

¹Mass Spectrometry Group, NIH/National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, United States ²Computational Chemistry Group, NIH/National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, United States ³Faculty of Dentistry, Srinakharinwirot University, Bangkok, Thailand ⁴School of Dentistry, University of North Carolina at Chapel Hill, North Carolina, United States ⁵Department of Bioengineering, Temple University, Pennsylvania, USA

Abstract

Collagens are the most abundant glycoproteins in the body. One characteristic of this protein family is that the amino acid sequence consists of repeats of three amino acids $-(X-Y-Gly)_n$. Within this motif, the Y residue is often 4-hydroxyproline (HyP) or 5-hydroxylysine (HyK). Glycosylation in collagen occurs at the 5-OH group in HyK in the form of two glycosides, galactosylhydroxylysine (Gal-HyK) and glucosyl galactosylhydroxylysine (GlcGal-HyK). In collision induced dissociation (CID), collagen tryptic glycopeptides exhibit unexpected gas-phase dissociation behavior compared to typical *N*- and *O*-linked glycopeptides, i.e. in addition to glycosidic bond cleavages, extensive cleavages of the amide bonds are observed. The Gal- or GlcGal- glycan modifications are largely retained on the fragment ions. These features enable unambiguous determination of the amino acid sequence of collagen glycopeptides and the location of the glycosylation site. This dissociation pattern was consistent for all analyzed collagen glycopeptides, regardless of their length or amino acid composition, collagen type or tissue. The two fragmentation pathways – amide bond and glycosidic bond cleavage – are highly competitive in collagen tryptic glycopeptides. The number of ionizing protons relative to the number of basic sites (i.e. Arg, Lys, HyK and N-terminus) is a major driving force of the fragmentation. We present here our experimental results and employ quantum mechanics calculations, to understand the factors enhancing the labile character of the amide bonds and the stability of hydroxylysine glycosides in gas phase dissociation of collagen glycopeptides.

Keywords

collagen; glycopeptide; hydroxyproline; galactosylhydroxylysine; tandem mass spectrometry; fragmentation pathway; mobile proton

Introduction

Glycosylation is recognized as one of the most biologically important post-translational protein modifications, being present on more than 50% of secreted proteins ¹, and evidence suggests that it is involved in a multitude of physiological events, such as protein trafficking,

*Corresponding authors: Irina Perdivara, Mass Spectrometry Group, NIH/National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, perdivarai@niehs.nih.gov, Phone: +1 (919) 541-0192, Fax: +1 (919) 541-0220. Kenneth B. Tomer, Mass Spectrometry Group, NIH/National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, tomer@niehs.nih.gov, Phone: +1 (919) 541-1966, Fax: +1 (919) 541-0220.

secretion, scaffolding, protein-protein and cell-cell interactions. Consequently, there has been an increasing interest/demand to characterize this modification at the molecular level, and mass spectrometry has become the technique of choice to address such tasks. Despite, however, recent advances in analytical approaches and instrumentation, characterization of glycosylated residues in proteins at the glycopeptide level still presents a major analytical challenge. This is due to several factors: (i) glycosylation is heterogeneous, a variety of glycoforms can be present at a specific site; (ii) sites of modification are difficult to predict, especially for *O*-linked structures, for which no consensus sequence has been defined; (iii) glycosylation often occurs non-stoichiometrically, and might be present at extremely low levels; (iv) glycopeptides ionize poorly compared to non-modified peptides, and, hence, the sensitivity of their detection in enzymatic mixtures might be greatly affected; (v) in collision induced dissociation (CID), the dominant (and sometimes exclusive) fragmentation pathway is the cleavage of the glycosidic bond. Many research groups have shown that some of these issues can be overcome by employing various strategies²⁻⁷.

Tandem mass spectrometry (MS/MS) represents a key technique for the determination of modification sites in glycopeptides. For the most part, CID and electron capture-/transfer-dissociation (ECD/ETD) provide complimentary information for glycopeptide characterization⁸⁻¹². In CID, fragmentation of protonated glycopeptides occurs mainly at the glycosidic bonds yielding sugar oxonium ions and/or fragment ions arising from sequential loss of monosaccharides from the non-reducing glycan end. In many cases, the protonated molecule corresponding to the non-glycosylated peptide may be the only information about the peptide sequence. Fragment ions arising from amide bond cleavage can sometimes be identified after significant magnification of the corresponding region(s) in the MS/MS spectrum^{4, 5, 13}. Often these low abundance ion can only be confidently used to verify the sequence of a known peptide. In contrast to mucin-type *O*-glycosylation, characteristic of the majority of secreted proteins, *O*-linked glycosylation of hydroxylysine residues is specific to the collagen family and to a small number of serum proteins containing collagenous sequence motifs, such as complement C1q, mannose binding lectin, and lung surfactant proteins A and D¹⁴.

The amino acid sequence of collagens is unique, consisting of repeats of three amino acids, $--(X-Y-Gly)_n--$. When *Y* is occupied by Lys or Pro, these residues can be hydroxylated by enzymes from the lysyl hydroxylase and prolyl hydroxylase families, resulting in formation of 5-hydroxylysine (HyK) and 4-hydroxyproline (HyP)^{15, 16}. The high content of Pro and HyP within the $--(X-Y-Gly)--$ repeats is essential for stabilization of the collagen triple helix through inter-chain hydrogen bonds. HyK residues can also be substrates for the collagen modifying enzymes, hydroxylysine galactosyl transferases and galactosylhydroxylysine glucosyl transferases, resulting in the formation of the *O*-glycosylated structures galactosylhydroxylysine (Gal-HyK) and glucosyl galactosylhydroxylysine (GlcGal-HyK)¹⁷⁻²⁰. The collagen proteins may contain varying amounts of Gal-HyK and GlcGal-HyK. The degree and pattern of collagen glycosylation varies depending on collagen types²¹, tissue types²²⁻²⁴ and pathological conditions^{25, 26}. For example, type I collagen is minimally glycosylated, containing, on average, one glycosylated HyK per 1000 residues, whereas type V collagen is heavily glycosylated, containing a tenfold higher extent of glycosylated HyK than type I²⁷.

Identification of the molecular loci of *O*-glycosylation in collagen was pioneered by Butler and Cunningham in the mid '60s²⁸ using a combination of proteolytic and chemical degradation, acid hydrolysis, paper chromatography and Edman sequencing. Modern approaches for collagen analysis are largely based on CNBr cleavage, size exclusion chromatography and mass spectrometry²⁹⁻³⁵. Recently, Taga *et al.* reported a selective

method for enrichment of collagen glycopeptides from a collagen tryptic digest using immobilized galactose oxidase, hydrazide chemistry and mass spectrometry ³⁶.

We previously employed liquid chromatography–tandem mass spectrometry (LC-MS/MS) to characterize the galactosylhydroxylysine glucosyl-transferase function of the enzyme lysyl hydroxylase-3 in the modification of type I collagen in mouse osteoblasts ³⁷. During this study, we observed that, when fragmented by CID, collagen-derived tryptic glycopeptides exhibit unexpected gas-phase dissociation behavior compared to non-collagen glycopeptides, i.e. in addition to glycosidic bond cleavages, peptide bonds undergo extensive cleavages in CID as well. Furthermore, fragment ions formed by peptide bond cleavage retain, for the most part, the glycan moiety attached to the hydroxyl group of HyK. The majority of the amide bond cleavages occurred N-terminal to HyP and Pro. This is not highly unusual, because there have been reports of high relative abundance of amide bond cleavages N-terminal to Pro ^{38, 39}. However, in the present example the specific cleavage of the 2–3 bond is a, or the, dominant fragmentation even though there are numerous HyP and Pro groups in the sequence of the collagen glycopeptides. Moreover, depending on charge state, fragmentation of the amide bonds and the glycosidic bonds become competitive. These atypical features enable amino acid sequence determination of the peptide, and identification of the glycosylation sites, such that additional fragmentation techniques (e.g. ETD) were found unnecessary. We present here our experimental results obtained in the MS/MS-based structural characterization of collagen glycopeptides. Furthermore, we employ these experimental observations as the basis for quantum mechanics calculations, to understand the factors enhancing the labile character of the amide bonds in gas phase dissociation of collagen glycopeptides.

Experimental

Materials and methods

Formic acid was from Sigma Aldrich (St Louis, MO); acetonitrile from Caledon Laboratories, Ltd. (Georgetown, Ontario); TPCK-modified porcine trypsin from Promega (Madison, WI). Deionized water was from an in-house Hydro Picopure 2 system (Durham, NC). All chemicals were used without further purification unless otherwise specified.

Collagen was isolated from cell cultures and various tissues (bovine bone, skin, and cartilage) as described previously ^{40, 41}, separated by SDS-PAGE, trypsinized and analyzed by LC-MS/MS to characterize the glycosylation sites ³⁷.

UPLC-nanoESI-QT of analyses were performed on a nanoAcquity UPLC - QToF Premier mass spectrometer (Waters Corporation, Milford, MA). The solvents were: 0.1% formic acid in deionized water (solvent A), and 0.1% formic acid in acetonitrile (solvent B). Following injection, the peptide mixture was trapped on a Symmetry C₁₈ column (5 μ m, 180 μ m $\times$ 50 mm, Waters, Milford, MA) at 2% solvent B, for 5 minutes at 5 μ L/min. Chromatography was performed on a C₁₈ Atlantis column (3 μ m, 100 μ m $\times$ 100 mm, Waters, Milford, MA), at 300 nL/min and separation was achieved with a linear gradient from 2 to 40% solvent B over 90 min. Mass spectrometric data were acquired using data dependent acquisition of the four most abundant ions, selected based on charge state criteria. For MS/MS a collision energy ramp from 20 to 40 V was employed. Instrument settings in the positive ion mode were as follows: capillary voltage: -- 3.5 kV, cone voltage 30 V, source temperature 80 C.

nanoHPLC-ESI-Ion trap analyses were performed on an Agilent 6340 ion trap (Santa Clara, CA) equipped with an HPLC ChipCube MS interface, an Agilent 1200 nanoHPLC and an electron transfer dissociation module, as described ¹⁰.

Quantum mechanical calculations were performed by employing Gaussian 09 at the wB97-XD level of theory with the 6-311++G** basis set for: (1) $\text{H}_3\text{N}^+\text{-Gly-Gly-Xxx-NMe}$, where Xxx is successively HyP, then Pro or Ala as controls, and NMe is the C-terminal N-methyl group, and (2) ${}^{\beta}\text{CH}_3\text{-}{}^{\gamma}\text{CH}_2\text{-}{}^{\delta}\text{CH}(\text{O-}\beta\text{-D-Gal})\text{-}{}^{\epsilon}\text{CH}_2\text{-N}^+\text{H}_3$. System (2) mimics the side chain of Gal-HyK, without the backbone moiety $\text{-NH-C}_\alpha\text{H-CO-}$. The labels β - ϵ indicate the position of each atom within the side chain of HyK. The initial protonation sites for each system were chosen based on the ‘mobile proton’ model^{42, 43}.

Results and Discussion

The notations α -1 and α -2 indicate the two chains of type I collagen. For example, α -1 [76–90] indicates peptide [76–90] of the α -1 chain. The following abbreviations will be used throughout the manuscript: HyP – (2S, 4R)-4-hydroxyproline, HyK – (2S, 5R)-5-hydroxy-L-lysine, Gal-HyK – galactosylhydroxylysine, GlcGal-HyK – glucosyl-galactosylhydroxylysine. In the discussion of the MS/MS data, the annotation ‘ y_n - GlcGal’ refers to the y_n fragment ion(s) which have lost the GlcGal glycan, while the annotation ‘ y_n ’ refers to the glycosylated fragment ion(s).

CID of collagen glycopeptides: fragmentation pathways in competition

To identify sites of glycosylation, a tryptic mixture of type I collagen was first analyzed by LC-MS/MS on a quadrupole ion trap. Using intensity based criteria for data dependent acquisition, collision induced dissociation spectra of the four most abundant ions were acquired¹⁰. The MS/MS of the ion of m/z 558.7 (3+), corresponding to α -1 [76–90] ${}^{76}\text{GLHyPGTAGLHyPGM}[(\text{Gal})\text{HyK}]\text{GHR}^{90}$ is shown in Figure 1A. Based on previous knowledge of collagen structure^{15, 16}, the following side chain modifications are consistent with the mass of this precursor ion: HyP 78, HyP 84, oxidized Met 86, and Gal-HyK 87. These modifications, however, cannot be unambiguously determined solely from this spectrum. The most abundant ions, observed as m/z 502.2 (3+) and 752.6 (2+), were assigned to glycosylated y_{13}^{3+} and y_{13}^{2+} . These arise from cleavage of the second amide bond from the N-terminus, between Leu 77 – HyP 78, i.e. the 2–3 amide bond. In addition to amide bond cleavage, ions arising from glycosidic bond cleavage were observed with lower relative abundance as m/z 448.2 (y_{13}^{3+} - Gal) and 671.6 (y_{13}^{2+} - Gal). The peptide [76–90] with GlcGal-HyK at residue 87 was observed as species of m/z 612.8 (3+) and 918.6 (2+), respectively. The ion trap CID spectra of these precursor ions are shown in Figures 1B and 1C. The identity and location of modifications cannot be fully determined, their assignment being based on known structural details of collagen^{15, 16}. Similar to the fragmentation of the 3+ charge state of the Gal-HyK containing glycopeptide [76–90] (Figure 1A), the major decomposition pathway in the MS/MS of m/z 612.8 (3+) corresponds to cleavage of the 2–3 amide bond, resulting in formation of the glycosylated y_{13} ions of m/z 556.3 (3+) and 833.6 (2+). In contrast, the MS/MS of the 2+ precursor ion of m/z 918.6, (Figure 1C), contains a large number of ions arising from both peptide- and glycosidic bond cleavages. Neutral loss of GlcGal represents the preferred fragmentation pathway in the MS/MS of the m/z 918.6 (2+) ion, because the fragment ion of m/z 756.6 (2+), assigned as the de-glycosylated doubly protonated peptide [76–90], represents the base peak in this spectrum. In addition, several C-terminal fragments showing loss of GlcGal were observed as ions of m/z 671.6 (y_{13}^{2+} - GlcGal), 1000.1 (y_9 - GlcGal), 830.1 (y_7 - GlcGal), 513.1 (y_4 - GlcGal). No y ions retaining the sugar moiety could be identified. Furthermore, neutral loss of 64 Da, corresponding to elimination of methanesulfenic acid (CH_3SOH) from the oxidized methionine side chain, was observed from the y -ions that have already lost the glycan, giving rise to additional fragments of m/z 639.6 (y_{13}^{2+} - GlcGal - CH_3SOH), 936.1 (y_9 - GlcGal - CH_3SOH), and 766.1 (y_7 - GlcGal - CH_3SOH).

Based on published literature, we expected that the dominant cleavage in the glycopeptide ions [76–90] would correspond to the glycosidic bond cleavage^{44–48}. Hence, amide bond scission(s) and retention of intact glycosidic bonds on the fragment ions were unexpected. Under higher energy collision conditions in the QToF system, the highest abundance fragment ion was m/z 1326.51 (y_{13} - GlcGal). The glycosylated ion y_{13} (m/z 1650.59) was the second most abundant. These data suggest that the initial fragmentation event is the cleavage of the 2–3 amide bond with retention of GlcGal on the y_{13} fragment ion. In the QToF, this fragment undergoes further fragmentation by sequential loss of the two monosaccharides, resulting in the base peak ion of m/z 1326.51 (Supplemental Figure S-1).

In fact, cleavage of the 2–3 amide bond between Leu 77 – HyP 78 in glycopeptides [76–90], was observed in the MS mode as well, based on the presence of the ion of m/z 556.0 (3+) in the MS scan. The perfect overlap of the extracted ion traces for the +3 ions of m/z 612.8 and 556.0 (data not shown) which differ by GL (residues 76 and 77) confirms that the latter is most likely a fragment arising from in-source decomposition of the former. The high relative abundance in the MS of the ion of m/z 556.0 (3+) due to the in source decay of m/z 612.8 (3+) triggered the selection of the former for MS/MS (Figure S-2). The major fragmentation event in this spectrum is the neutral loss of GlcGal from the precursor ion, m/z 556.0 (3+) → 671.6 (2+).

The MS/MS spectra shown in Figure 1 indicate that: (i) in the MS/MS of the 3+ precursor ions, the cleavage rate of the 2–3 amide bond is higher than that of the glycosidic bonds; (ii) the MS/MS spectra of the 3+ and 2+ precursor ions differ in the fragment ions observed; (iii) the 3+ glycoforms, Gal- and GlcGal-HyK, of the peptide [76–90] fragment similarly in CID, i.e. the 2–3 amide bond cleavage prevails over the glycosidic bond cleavage. This suggests that the length of the glycan does not significantly influence the dissociation of the corresponding precursor ions.

It is noteworthy that two amide bonds between identical amino acids and N-terminal to HyP exist in peptide [76–90], i.e. Leu 77 – HyP 78 (y_{13}), and Leu 83 – HyP 84 (y_7). The amide bond Leu 77 – HyP 78 represents, by far, the preferred fragmentation pathway in the CID spectrum of the 3+ molecular ion, because cleavage between Leu 83 – HyP 84 was observed only at elevated collision energies (Figure S-1). Of the MS/MS spectra discussed so far, the spectrum in Figure S-1B is the most informative for peptide sequence determination. However, this spectrum resembles that of the non-glycosylated peptide ion, and, with no identifiable diagnostic sugar ions, it is difficult to recognize this as the spectrum of a glycosylated species. A database search of this spectrum results in a false-positive assignment. This example of glycopeptide fragmentation in the ion trap and QToF is similar to a published literature example of a mucine type *O*-glycopeptide⁴⁹ where on the ion trap, in addition to the dominant glycosidic bond cleavages, some backbone fragmentation was observed. The major difference between this and our data is that, in our case, the dominant cleavage is the 2–3 HyP cleavage, and not the glycosidic bond.

Other examples of tryptic collagen glycopeptides CID spectra

Studies of typical *N*- and *O*-linked glycosylated peptides have shown that, under certain conditions, peptide sequence can be derived from CID^{45, 49, 50}. Key factors for the observation of backbone fragments include the nature of the charge carrier (H^+ vs. Na^+/Li^+ adducts), the number of ionizing protons, peptide length and glycan type (high-mannose vs. complex-type)⁴⁶. To determine the extent and driving factors of amide bond fragmentation in collagen glycopeptides, we analyzed a number of QToF-MS/MS spectra obtained at defined collision energies. Different charge states of the same peptide(s) were observed to undergo significantly different fragmentations. These are exemplified in Figure 2A–F comparing the MS/MS spectra of the 3+ and 4+ ions (see discussion below).

The deconvoluted CID spectra of m/z 1424.941 (3+) and 1068.978 (4+), obtained with a collision energy ramp of 20 – 30 V, were assigned to bovine type II collagen α -1 [193–237] containing Gal-HyK at residue 219 and eight HyPs (Figures 2A and 2B). In Figure 2A (note the magnification factor of 82), the main fragmentation pathway is the loss of hexose from the protonated molecule, based on the ion of m/z 4110.84. Another abundant ion, m/z 2913.43, could not be assigned to any predicted, glycosylated or non-glycosylated, sequence ion. Only two glycosylated sequence ions, y_{24} (m/z 2464.19) and y_{43} (m/z 4086.85) were observed. In contrast, the spectrum of the 4+ precursor shows abundant backbone fragmentation while the parent ion (m/z 4272.89) is not entirely fragmented. The major finding here is that all backbone fragments containing the site HyK 219, i.e. y_{22} , b_{32} , b_{36} , b_{38} and b_{41} , were observed in glycosylated form (Figure 2B). The glycosidic bond cleavage is minor, as suggested by the low abundance of the non-glycosylated ions, indicated with arrows in Figure 2B. Compared to Figure 2A, the glycan loss from the precursor ion (m/z 4272.89 $\rightarrow$ 4110.84) is no longer the dominant cleavage.

Two additional glycopeptides, α -1 [91–126] from bovine type II collagen (Figures 2C and 2D) and α -2 [193–237] from bovine type I collagen (Figures 2E and 2F), were characterized from the MS/MS of their 3+ and 4+ precursor ions. Type II peptide α -1 [91–126] contains two potential glycosylation sites, HyK 99 and HyK 108, and four HyPs. The CID spectrum of m/z 1239.21 (3+) shows two neutral losses of 162 Da each, suggesting the presence of two hexoses (Figure 2C). These are the only observed fragmentation events in this spectrum. The site(s) and glycan assignment (Gal- and Gal- vs. GlcGal-) are not possible. However, the spectrum of the 4+ precursor ion (Figure 2D) contains extensive sequence ions. The preferred dissociation channel is the cleavage of the 2–3 amide bond, with formation of the diglycosylated ion y_{34} (m/z 3495.51). Neutral loss of hexose(s) from the molecular ion occurs at a lower rate, because the corresponding non-glycosylated ions (m/z 3553.54 – loss of one hexose, and m/z 3391.53 – loss of two hexoses) have lower abundance compared to diglycosylated y_{34} . The glycosylation pattern was determined as Gal-HyK 99 and Gal-HyK 108 based on the presence of the monoglycosylated ions b_{14} (m/z 1438.63) and y_{22} (m/z 2277.96). The high relative abundance of the non-glycosylated ion y_{22} (m/z 2115.91) and the low abundance of the diglycosylated b_{14} ion (m/z 1600.65, not labeled in Figure 2D) suggests that the glycoform distribution GlcGal-HyK 99 and HyK 108 is also present on glycopeptide α -1 [91–126].

The MS/MS spectrum of m/z 1384.633 (3+) (Figure 2E) suggests that the dominant fragmentation is the loss of galactose from the triply protonated molecule (m/z 4151.90 $\rightarrow$ 3989.80). A significant number of glycosylated y ions were observed, few of which had relative abundances comparable to that originating from neutral loss of hexose from the precursor ion. In contrast, the spectrum of the 4+ precursor (Figure 2F) contains mainly de-glycosylated fragment ions and only a few lower abundance glycosylated species, suggesting that the former arose from the latter via multiple collisions.

Can a rule be derived for CID fragmentation of tryptic collagen glycopeptides?

The examples in Figures 1 and 2 suggest that precursor ions in the lower charge state are dominated by glycosidic bond cleavage, whereas those in higher charge states by amide bond cleavage. According to the mobile proton model^{42, 43}, protons are classified as mobile (number of protons > number of Arg, Lys, His and N-terminus), partially mobile (number of basic sites – number of protons > number of Arg) and non-mobile (number of protons – number of Arg). Glycopeptides α -1 [76–90] (Figure 1) are partially mobile, and those in Figure 2 are either ‘fully mobile’ in both 3+ and 4+ charges (Figures 2A–E and 2F) or ‘partially mobile’ (Figure 2C). Based on this classification, however, it appears that no correlation exists between proton mobilities and preferred dissociation pathways in CID.

The results can be partially rationalized by considering the number of ionizing protons relative to the total number of Arg, Lys, HyK and N-terminus. It appears that the amide bond cleavage is preferred when the number of protons is *greater* than the number of Arg, Lys, HyK and N-terminus, as demonstrated by the examples in Figures 2B and 2F. When the number of protons *equals* the number of sites, as is the case in Figures 1A–B, 2A, 2D–E, and S-2, amide- and glycosidic bond cleavages compete for priority. In some examples the amide bond clearly dominates (Figures 1A–B, 2D), whereas in others the glycosidic bond is preferred (Figures 2A, 2E). This suggests that in these cases other factors also contribute. When the number of protons is *less* than the number of Arg, Lys, Hyl and N-terminus (Figures 1C, 2C), glycosidic bond cleavages are almost exclusively observed. This observation is further supported by two examples shown in Figure 3. The MS/MS of m/z 851.143 (4+) (Figure 3A) corresponds to mouse type I collagen α -1 [556–585] with four hexoses on two potential glycosylation sites, HyK 564 and HyK 573, and four ionizing protons on six sites, i.e. N-terminus, two Arg, two Hyl and one Lys. The spectrum resembles that of a typical mucine *O*-glycopeptide, because the sequential losses of hexose from the protonated molecular ion dominate, whereas amide bond cleavages are minimal. The assignment was made based on the MS/MS spectrum of another glycoform showing more backbone fragmentation. The MS/MS in Figure 3B was assigned to α -1 [556–573] with two hexoses and one glycosylation site, HyK 564. The large number of non-glycosylated backbone fragments and the non-glycosylated molecular ion suggest that the glycan loss is the first decomposition event. Three protons on four sites are present, i.e. N-terminus, one Arg, one Hyl and one Lys. The spectrum resembles that of the non-glycosylated peptide, rendering automatic glycosylation site assignment impossible. Another example of a glycopeptide with four hexoses on two glycosites is shown in Figure S-3.

Hence, following modification of the mobile proton rules can be made for fragmentation of collagen tryptic glycopeptides: amide- and glycosidic bond cleavages compete with each other for dissociation in glycopeptides containing equal number of ionizing protons to the number of sites, whereas in those glycopeptides where the number of protons is greater, the amide bond is preferentially cleaved.

In CID, the class of collagen *O*-glycopeptides exhibits unique characteristics compared to mucine type *O*-glycopeptides, because their rate of amide bond cleavage can be higher than that of the glycosidic bonds. However, a direct comparison between the two classes cannot be made, because of the differences in the nature of sugar acceptor (HyK vs. Ser/Thr) and the attached sugars (Gal vs. GalNAc). It is noteworthy that amide bond cleavages in mucine *O*-glycopeptides have been reported numerous times and sequencing was possible with location of the glycosylation sites^{4, 5, 13, 49, 51}. The main difference between these published examples and collagen glycopeptides is that in mucine *O*-glycopeptides neutral loss of the sugars from the precursor ion represents, almost exclusively, the main dissociation channel.

Quantum mechanics calculations (QM)

To understand the unusual labile character of the 2–3 amide bond in glycopeptides [76–90], model tripeptides Gly-Gly-Xxx-NMe were employed for quantum mechanics calculations. Xxx was chosen as (2S, 4R)-4-hydroxyproline (HyP), and then Pro and Ala (as controls for cyclic vs. alkyl side chains). One ionizing proton was initially assumed at the N-terminus of Gly-Gly-Xxx⁴³. All possible conformers were optimized for each of the three N-terminally protonated systems, and various optimized conformations were subjected to mobilization of the ionizing proton from the N-terminus to the second amide nitrogen, as a first step in the fragmentation process⁴³. One of the lowest energy structures of H_3^+N -Gly-Gly-HyP-NMe is shown in Figure 4A, with the closest hydrogen of the H_3N^+ group being 2.9 Å away from the HyP amide nitrogen. Two hydrogen bonds of the N-terminus protons, one with the 4-OH group of HyP (1.98 Å) and one with the second amide oxygen (1.86 Å), stabilize this

structure. In the optimized structures when Xxx is Pro and Ala, the closest distances between any N-terminal proton and the amide nitrogen of Xxx were found 3.3 Å and 3.9 Å, respectively. The energy profile for the proton transfer from the N-terminus to the amide nitrogen of HyP is shown in Figure 4B. The formation of a metastable H-bond between the protonated N-terminus and the nitrogen of the second amide bond is suggested by the plateau in the middle portion of the graph. No additional activation energy is needed for proton transfer. The energies required for proton transfer to the Xxx amide nitrogen were 15.9, 26.3 and 36.3 kcal/mol for HyP, Pro and Ala, respectively.

The energy required to separate the two fragments at the amide bond Gly-Xxx was estimated using single point energies of the system at various distances of the Gly carbonyl carbon and Xxx amide nitrogen (Figure S-4). Note that the starting conformation is energy optimized, but the resulting fragments at various distances are frozen in their initial conformation. The calculations were performed for systems with the ionizing proton either on the N-terminus or on the Xxx amide nitrogen. In the first case, the energies required for fragmentation of the amide bond Gly-Xxx were 184.4, 200.2, and 198.3 kcal/mol for Xxx=HyP, Pro, and Ala, respectively. However, when the ionizing proton was mobilized at the Xxx amide nitrogen, the fragmentation energies were substantially lower, i.e. 96.6, 97.9, and 109.2 kcal/mol for Xxx= HyP, Pro, and Ala, respectively. This implies that the fragmentation pathway involving mobilization of the ionizing proton to the Xxx amide nitrogen is energetically favored compared to the pathway when the ionizing proton is at the N-terminus, i.e. (15.9+96.6) vs. 184.4 kcal/mol for HyP, (26.3+97.9) vs. 200.2 kcal/mol for Pro, and (36.3+109.2) vs. 198.3 kcal/mol for Ala. In summary, our results support that the mobile proton mechanism facilitates the breaking of 2–3 amide bond and that the ‘ease of fragmentation’ is Hyp > Pro > Ala.

The unusual stability of the glycosidic bond in the hydroxylysine glycosides Gal-HyK and GlcGal-HyK is suggested by the MS/MS spectra shown in Figures 1 and 2. A model system, $\beta\text{CH}_3\text{-}\gamma\text{CH}_2\text{-}\delta\text{CH}(\text{O-}\beta\text{-D-Gal})\text{-}\epsilon\text{CH}_2\text{-N}^+\text{H}_3$ containing the side chain of hydroxylysine between $\beta\text{C-}\epsilon\text{C}$ (i.e. without the backbone moiety $\text{-NH-C}\alpha\text{H-CO-}$) was used to determine how much energy is required to fragment the glycosidic bond. Initial protonation was assumed at the ϵ -amino group⁴² with a galactosylated 5-OH group and a D-galactopyranose ring in a β -configuration. The β -CH₂ group was capped with a hydrogen atom. The lowest energy conformation determined for this system is shown in Figure 4C. Two hydrogen bonds characterize this system, one between the proton of the $\epsilon\text{-N}^+\text{H}_3$ group and the OH group at C6 in the galactopyranose ring (1.77 Å), and the second one with the glycosidic oxygen (2.09 Å). After proton transfer to the glycosidic oxygen, the re-optimized new structure (Figure 4D) was found 38.5 kcal/mol less stable than the initial structure (Figure 4C). This value represents the minimal amount of energy necessary to mobilize the ionizing proton from the $\epsilon\text{-N}^+\text{H}_3$ group to the glycosidic oxygen. Next, fragmentation energies at the glycosidic bond in the systems shown in Figures 4C and 4D were determined. The system in Figure 4C requires 198.3 kcal/mol for fragmentation, while the system in Figure 4D requires only 78.9 kcal/mol. When combining the proton transfer- and the fragmentation energies, the resulting energy value of 117.4 kcal/mol is comparable to the calculated energies required for fragmenting the amide bond (112.5 kcal/mol for HyP and 145.5 kcal/mol for Ala). However, this value underestimates the true fragmentation energy of the glycosidic bond, because the interactions of Gal with the peptide backbone were not considered in this simplified model. Due to the large number of possible configurations, the sampling is time consuming. However, we analyzed one such configuration and we found an additional 15 kcal/mol required for fragmentation. This suggest that, under ‘mobile proton’ conditions, cleavage of the amide bond between Leu 77 – Hyp 78 in glycopeptides [76–90] is probably favored over cleavage of the glycosidic bond in Gal-HyK. If the proton were mobilized to the glycosidic oxygen from another protonation site in the molecule, this would

require coexistence of the $\epsilon\text{-N}^+\text{H}_3$ group and the protonated glycosidic oxygen. This structure, however, is 97 kcal/mol less stable than the initial one, suggesting that population of this channel is highly unlikely. These combined results are further supported by an early publication reporting that the glycosides of hydroxylysine represent the final products of collagen degradation in humans⁵². This suggests that Gal-HyK and GlcGal-HyK are extremely stable under physiologic conditions, and that they can be used as biological markers of bone resorption²².

Conclusion

In conclusion, we combined experimental results and theoretical studies to explain some remarkable features observed in the MS/MS spectra of tryptic collagen glycopeptides. The energy values determined using quantum mechanics calculations support our interpretation of the MS/MS data and proved helpful in understanding the gas-phase dissociation behavior of protonated collagen glycopeptides.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (ES050171 to KBT and ES043010 to LP), and by the NIH grants AR060978 (MY) and R01 AR056145 (NP). The authors would like to acknowledge Dr. Joshua Sharp (CCRC, Athens, GA) and Dr. Lee Pedersen (NIEHS/NIH) for the critical review of the manuscript, and the NIEHS Protein Microcharacterization Core Facility for the assistance with the in-gel digests.

References

1. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J. Essentials of Glycobiology. 2009
2. Settineri CA, Medzihradszky KF, Masiarz FR, Burlingame AL, Chu C, George-Nascimento C. Characterization of O-glycosylation sites in recombinant B-chain of platelet-derived growth factor expressed in yeast using liquid secondary ion mass spectrometry, tandem mass spectrometry and Edman sequence analysis. *Biomed Environ Mass Spectrom*. 1990; 19(11):665–76. [PubMed: 2076465]
3. Medzihradszky KF, Gillece-Castro BL, Hardy MR, Townsend RR, Burlingame AL. Structural Elucidation of O-linked Glycopeptides by High Energy Collision-Induced Dissociation. *Journal of the American Society for Mass Spectrometry*. 1996; 7:319–328.
4. Hanisch FG, Green BN, Bateman R, Peter-Katalinic J. Localization of O-glycosylation sites of MUC1 tandem repeats by QTOF ESI mass spectrometry. *J Mass Spectrom*. 1998; 33(4):358–62. [PubMed: 9597769]
5. Alving K, Paulsen H, Peter-Katalinic J. Characterization of O-glycosylation sites in MUC2 glycopeptides by nanoelectrospray QTOF mass spectrometry. *J Mass Spectrom*. 1999; 34(4):395–407. [PubMed: 10226365]
6. Calvano CD, Zamboni CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *Journal of Proteomics*. 2008; 71(3):304–317. [PubMed: 18638581]
7. Wuhler M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007; 849(1–2): 115–28.
8. Halim A, Nilsson J, Ruetschi U, Hesse C, Larson G. Human urinary glycoproteomics: attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics*. 11(4):M111 013649. [PubMed: 22171320]

9. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 2004; 101(26):9528–33. [PubMed: 15210983]
10. Perdivara I, Petrovich R, Allinquant B, Deterding LJ, Tomer KB, Przybylski M. Elucidation of O-glycosylation structures of the beta-amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision induced dissociation. *J Proteome Res*. 2009; 8(2):631–42. [PubMed: 19093876]
11. Peter-Katalinic J. Methods in enzymology: O-glycosylation of proteins. *Methods Enzymol*. 2005; 405:139–71. [PubMed: 16413314]
12. Medzihradsky KF. Peptide sequence analysis. 2005; 402:209–44.
13. Chalkley RJ, Burlingame AL. Identification of novel sites of O-N-acetylglucosamine modification of serum response factor using quadrupole time-of-flight mass spectrometry. *Mol Cell Proteomics*. 2003; 2(3):182–90. [PubMed: 12684542]
14. Kolble K, Reid KB. The genomics of soluble proteins with collagenous domains: C1q, MBL, SP-A, SP-D, conglutinin, and CL-43. *Behring Inst Mitt*. 1993; (93):81–6. [PubMed: 8172588]
15. Kivirikko KI, Myllyla R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol*. 1982; 82(Pt A):245–304. [PubMed: 6210830]
16. Kivirikko KI, Myllyla R. Recent developments in posttranslational modification: intracellular processing. *Methods Enzymol*. 1987; 144:96–114. [PubMed: 3041180]
17. Liefhebber JM, Punt S, Spaan WJ, van Leeuwen HC. The human collagen beta(1-O)galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. *BMC Cell Biol*. 11:33. [PubMed: 20470363]
18. Wang C, Luosjarvi H, Heikkinen J, Risteli M, Uitto L, Myllyla R. The third activity for lysyl hydroxylase 3: galactosylation of hydroxyllysyl residues in collagens in vitro. *Matrix Biol*. 2002; 21(7):559–66. [PubMed: 12475640]
19. Myllyla R, Wang C, Heikkinen J, Juffer A, Lampela O, Risteli M, Ruotsalainen H, Salo A, Sipila L. Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J Cell Physiol*. 2007; 212(2):323–9. [PubMed: 17516569]
20. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennen T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. *Mol Cell Biol*. 2009; 29(4):943–52. [PubMed: 19075007]
21. Shinkai H, Yonemasu K. Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. *Biochem J*. 1979; 177(3):847–52. [PubMed: 444207]
22. Segrest JP, Cunningham LW. Variations in human urinary O-hydroxyllysyl glycoside levels and their relationship to collagen metabolism. *J Clin Invest*. 1970; 49(8):1497–509. [PubMed: 5431661]
23. Schofield JD, Freeman IL, Jackson DS. The isolation, and amino acid and carbohydrate composition, of polymeric collagens prepared from various human tissues. *Biochem J*. 1971; 124(3):467–73. [PubMed: 5002672]
24. Pinnell SR, Fox R, Krane SM. Human collagens: differences in glycosylated hydroxylsines in skin and bone. *Biochim Biophys Acta*. 1971; 229(1):119–22. [PubMed: 5101159]
25. Tenni R, Valli M, Rossi A, Cetta G. Possible role of overglycosylation in the type I collagen triple helical domain in the molecular pathogenesis of osteogenesis imperfecta. *Am J Med Genet*. 1993; 45(2):252–6. [PubMed: 8456811]
26. Dominguez LJ, Barbagallo M, Moro L. Collagen overglycosylation: a biochemical feature that may contribute to bone quality. *Biochem Biophys Res Commun*. 2005; 330(1):1–4. [PubMed: 15781223]
27. Mizuno K, Adachi E, Imamura Y, Katsumata O, Hayashi T. The fibril structure of type V collagen triple-helical domain. *Micron*. 2001; 32(3):317–23. [PubMed: 11006511]
28. Butler WT, Cunningham LW. Evidence for the linkage of a disaccharide to hydroxylysine in tropocollagen. *J Biol Chem*. 1966; 241(17):3882–8. [PubMed: 4288358]
29. Henkel W, Dreisewerd K. Cyanogen bromide peptides of the fibrillar collagens I, III, and V and their mass spectrometric characterization: detection of linear peptides, peptide glycosylation, and

- cross-linking peptides involved in formation of homo- and heterotypic fibrils. *J Proteome Res.* 2007; 6(11):4269–89. [PubMed: 17939700]
30. Hanson DA, Eyre DR. Molecular site specificity of pyridinoline and pyrrole cross-links in type I collagen of human bone. *J Biol Chem.* 1996; 271(43):26508–16. [PubMed: 8900119]
 31. Eyre DR, Weis MA, Wu JJ. Maturation of collagen Ketoimine cross-links by an alternative mechanism to pyridinoline formation in cartilage. *J Biol Chem.* 285(22):16675–82. [PubMed: 20363745]
 32. Eyre DR, Pietka T, Weis MA, Wu JJ. Covalent cross-linking of the NC1 domain of collagen type IX to collagen type II in cartilage. *J Biol Chem.* 2004; 279(4):2568–74. [PubMed: 14602708]
 33. Mattu TS, Royle L, Langridge J, Wormald MR, Van den Steen PE, Van Damme J, Opdenakker G, Harvey DJ, Dwek RA, Rudd PM. O-glycan analysis of natural human neutrophil gelatinase B using a combination of normal phase-HPLC and online tandem mass spectrometry: implications for the domain organization of the enzyme. *Biochemistry.* 2000; 39(51):15695–704. [PubMed: 11123894]
 34. Van den Steen PE, Proost P, Brand DD, Kang AH, Van Damme J, Opdenakker G. Generation of glycosylated remnant epitopes from human collagen type II by gelatinase B. *Biochemistry.* 2004; 43(33):10809–10816. [PubMed: 15311942]
 35. Yang CX, Park AC, Davis NA, Russell JD, Kim B, Brand DD, Lawrence MJ, Ge Y, Westphall MS, Coon JJ, Greenspan DS. Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid residues of the alpha 1(V) Collagen Chain. *Journal of Biological Chemistry.* 2012; 287(48):40598–40610. [PubMed: 23060441]
 36. Taga Y, Kusubata M, Ogawa-Goto K, Hattori S. Development of a Novel Method for Analyzing Collagen O-glycosylations by Hydrazide Chemistry. *Mol Cell Proteomics.* 2012; 11(6):M111010397. [PubMed: 22247541]
 37. Sricholpech M, Perdivara I, Yokoyama M, Nagaoka H, Terajima M, Tomer KB, Yamauchi M. Lysyl Hydroxylase 3-mediated Glucosylation in Type I Collagen: MOLECULAR LOCATIONS AND BIOLOGICAL SIGNIFICANCE. *J Biol Chem.* 2012; 287(27):22998–3009. [PubMed: 22573318]
 38. Brei LA, Tabb DL, Yates JR 3rd, Wysocki VH. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem.* 2003; 75(9):1963–71. [PubMed: 12720328]
 39. Huang Y, Tseng GC, Yuan S, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. A data-mining scheme for identifying peptide structural motifs responsible for different MS/MS fragmentation intensity patterns. *J Proteome Res.* 2008; 7(1):70–9. [PubMed: 18052120]
 40. Yamauchi M, Shiiba M. Lysine hydroxylation and cross-linking of collagen. *Methods Mol Biol.* 2008; 446:95–108. [PubMed: 18373252]
 41. Sricholpech M, Perdivara I, Nagaoka H, Yokoyama M, Tomer KB, Yamauchi M. Lysyl hydroxylase 3 glucosylates galactosylhydroxyllysine residues in type I collagen in osteoblast culture. *J Biol Chem.* 2011; 286(11):8846–56. [PubMed: 21220425]
 42. Wysocki VH, Tsaprailis G, Smith LL, Brei LA. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom.* 2000; 35(12):1399–406. [PubMed: 11180630]
 43. Paizs B, Suhai S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev.* 2005; 24(4):508–48. [PubMed: 15389847]
 44. Domon B, Costello CE. A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates. *Glycoconjugate J.* 1988; 5:397–409.
 45. Iacob R, Perdivara I, Przybylski M, Tomer KB. Mass spectrometric characterization of glycosylation of Hepatitis C virus E2 envelope glycoprotein reveals extended microheterogeneity of N-glycans. *JASMS.* 2007
 46. Dodds ED. Gas-phase dissociation of glycosylated peptide ions. *Mass Spectrom Rev.* 2005
 47. Wührer M, Stam JC, van de Geijn FE, Koeleman CA, Verris CT, Dolhain RJ, Hokke CH, Deelder AM. Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics.* 2007; 7(22):4070–81. [PubMed: 17994628]
 48. Wührer M, Deelder AM, Hokke CH. Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2005; 825(2):124–33.

49. Darula Z, Chalkley RJ, Lynn A, Baker PR, Medzihradszky KF. Improved identification of O-linked glycopeptides from ETD data with optimized scoring for different charge states and cleavage specificities. *Amino Acids*. 2011; 41(2):321–8. [PubMed: 20652609]
50. Darula Z, Medzihradszky KF. Affinity Enrichment and Characterization of Mucin Core-1 Type Glycopeptides from Bovine Serum. *Molecular & Cellular Proteomics*. 2009; 8(11):2515–2526. [PubMed: 19674964]
51. Darula Z, Medzihradszky KF. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics*. 2009; 8(11):2515–26. [PubMed: 19674964]
52. Moro L, Norissuarez K, Michalsky M, Romanello M, Debernard B. The Glycosides of Hydroxylysine Are Final Products of Collagen Degradation in Humans. *Biochimica Et Biophysica Acta*. 1993; 1156(3):288–290. [PubMed: 8461318]

Figure 1.

Ion trap CID of glycopeptides α -1 [76–90], for the precursor ions: **A.** m/z 558.7 (3+), **B.** m/z 612.7 (3+), and **C.** m/z 918.6 (2+). Each glycopeptide contains oxidized Met (M_{ox}), two HyP and either Gal-HyK or GlcGal-HyK, as indicated. Arrows indicate loss of CH_3SOH ($-64 Da$) from M_{ox} .

Figure 2. Deconvoluted QToF CID spectra of the 3+ and 4+ precursor ions of three glycopeptides: **A.** and **B.** – bovine type II collagen α -1 [193–237]; **C.** and **D.** – bovine type II collagen α -1 [91–126]; **E.** and **F.** – bovine type I collagen α -1 [193–237]. Following collision energy ramps were employed: 20 – 30 V for the spectra in **A.**, **B.** and **D.**, and 30 – 40 V for the spectra in **C.**, **E.** and **F.**

Figure 3. Deconvoluted QToF CID spectra of glycopeptide ions of: **A.** m/z 851.143 (4+), corresponding to α -1 [556–585], and **B.** m/z 659.663 (3+), corresponding to α -1 [556–573] from mouse type I collagen. The spectra were obtained with a collision energy ramp of 30 – 40 V (**A.**), and 25 – 30 V (**B.**).

Figure 4.

A. One lowest energy structure of H_3N^+ -Gly-Gly-HyP-NMe determined at the wb97-XD level of Gaussian calculations with the 6-311++G** basis set; **B.** Energy profile for the proton transfer from protonated N-terminus to the second amide nitrogen in Gly-Gly-HyP; **C.** Lowest energy structure determined for the system ${}^{\beta}\text{CH}_3$ - ${}^{\gamma}\text{CH}_2$ - ${}^{\delta}\text{CH}(\text{O}-\beta\text{-D-Gal})$ - ${}^{\epsilon}\text{CH}_2$ - N^+H_3 when the ionizing proton is at the ϵ -amino group, and **D.** when the ionizing proton was transferred to the glycosidic oxygen.