Published in final edited form as:
Aust J Chem. 2013 July 18; 66(7): 760–769. doi:10.1071/CH13174.

Molecular Characterization of Collagen Hydroxylysine *O*-Glycosylation by Mass Spectrometry: Current Status

Irina Perdivara[A,C], **Mitsuo Yamauchi**[B], and **Kenneth B. Tomer**[A]

[A]Mass Spectrometry Group, National Institutes of Health/National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, NC 27709, USA.

[B]School of Dentistry, University of North Carolina at Chapel Hill, North Carolina, NC 27599, USA.

Abstract

The most abundant proteins in vertebrates – the collagen family proteins – play structural and biological roles in the body. The predominant member, type I collagen, provides tissues and organs with structure and connectivity. This protein has several unique post-translational modifications that take place intra- and extra-cellularly. With growing evidence of the relevance of such post-translational modifications in health and disease, the biological significance of *O*-linked collagen glycosylation has recently drawn increased attention. However, several aspects of this unique modification – the requirement for prior lysyl hydroxylation as a substrate, involvement of at least two distinct glycosyl transferases, its involvement in intermolecular crosslinking – have made its molecular mapping and quantitative characterization challenging. Such characterization is obviously crucial for understanding its biological significance. Recent progress in mass spectrometry has provided an unprecedented opportunity for this type of analysis. This review summarizes recent advances in the area of *O*-glycosylation of fibrillar collagens and their characterization using state-of-the-art liquid chromatography–mass spectrometry-based methodologies, and perspectives on future research. The analytical characterization of collagen crosslinking and advanced glycation end-products are not addressed here.

Introduction

Collagens represent a large family of extracellular matrix proteins comprising at least 29 different genetic types encoded by 44 genes.[1] These are the most abundant proteins in vertebrates, and collagens are classified into several sub-family groups, depending on the molecular structure and assembly mode.[2] Of the sub-families, the fibril-forming group, including types I, II, III, V, XI, XXIV, and XXVII, represents the largest and most characterized, except for the latter two types that were recently identified. Type I collagen is the ‘classical’ type and is the main structural component in almost all connective tissues such as skin, tendon, bone, and ligament. In addition to the structural role, it also regulates cellular activity by interacting with their specific cell surface receptors such as integrin,

[C]Corresponding author. perdivarai@niehs.nih.gov.

discoidin domain receptors.[3] The hallmark of collagen is the unique triple helical structure formed by intertwining three polypeptides, 'α' chains (Fig. 1a). The amino acid sequences of the α chains in the triple helix consist of –(Gly–X–Y)$_n$– triplet repeats. The presence of Gly, the smallest amino acid, at every third residue is a prerequisite for the formation of the triple helix as it is the only amino acid that can physically fit in the centre of the collagen helix. Another unique characteristic of collagen is the presence of post-translationally modified amino acids at position Y, including 4-hydroxyproline (Hyp), 5-hydroxylysine (Hyl), and glycosylated Hyl[4,5] (Fig. 1a), and at position X, of 3-hydroxyproline (3-Hyp). The high content of Hyp is required for the stability of the triple-helical structure, whereas Hyl is crucial for collagen glycosylation and crosslinking. The helical 'collagenous domain' –(Gly–X–Y)$_n$ and the associated modifications are also found in several collagen-like proteins, such as mannose-binding lectin (MBL), complement C1q, adiponectin, and lung surfactant proteins SP-A and SP-D.[6]

Collagen Glycosylation: Structure and Function

Glycosylation of collagen was first identified in 1935,[7] but the structures of the glycosides were not defined until the mid-1960s.[8,9] Collagen biosynthesis and post-translational modifications of the nascent chains, including hydroxylation and glycosylation, take place in the endoplasmic reticulum (ER), before assembly into the triple helix to form a procollagen molecule in cells. Collagen *O*-glycosylation is first formed by the attachment of galactose (Gal) in its β-configuration to the 5-hydroxyl group of Hyl to form galactosylhydroxylysine (Gal–Hyl). This step is catalyzed by two newly identified, ER-resident enzymes, Glt25d1 and Glt25d2.[10,11] The second step is the transfer of glucose (Glc) to the Gal–Hyl residues, likely by a multifunctional collagen-modifying enzyme lysyl hydroxylase 3 (LH3), to form glucosyl galactosylhydroxylsine (GlcGal–Hyl) (Fig. 1b).[12–14] These are the only two *O*-linked glycosides found in mature type I collagen. Therefore, the main difference between collagen- and mucin-type *O*-glycosylation is the nature of the sugar acceptor (Hyl versus Ser and Thr) and the nature of the attached sugar (Gal versus GlcNAc).

The attachment sites, pattern (mono- or diglycosylation), and number of glycosylated structures differ among collagen types. For example, type I collagen is the least glycosylated member, containing on average one Gal–Hyl and one GlcGal–Hyl residue every 1000 amino acids, i.e. ~two glycosylated Hyl residues per α chain. In contrast, type IV and V collagen are heavily glycosylated, containing 2 Gal–Hyl and up to 44 GlcGal–Hyl residues per 1000 amino acids (type IV) and up to 7 Gal-Hyl and 44 GlcGal–Hyl residues (type V) per 1000 amino acids.[15,16] However, even within the same type of collagen, the amount, type, and distribution of glycosylation appear to depend on the tissue, cell type and pathophysiologic condition. For example, type I collagen is differentially glycosylated depending on functional regions within tissue,[17] maturation stage,[18] and disease state.[19–27]

The functional role of collagen glycosylation is not clearly defined as yet. Some studies suggest that glycosylation increases collagen stability against proteolytic degradation, controls the lateral growth of fibrils, and mediates interaction with noncollagenous molecules.[28–32] Some studies have also indicated that glycosylation in fibrillar collagen functions as a ligand for collagen-specific cell-surface receptors, e.g. discoidin domain

receptors (DDR) 1 and 2,[29–32] thus controlling certain signalling pathways. During tumour cell invasion, a critical step in the onset of metastasis, melanoma cell receptors bind to triple-helical regions of network-forming type IV collagen present in the basement membranes.[33] Receptor binding is greatly diminished when glycosylation is present in the key regions of type IV collagen molecules, minimizing tumour cell invasion. As another example, in autoimmune rheumatoid arthritis, glycosylated epitopes derived from type II collagen in synovial fluid may be critical in mediating the inflammatory process, by recruiting T-cells to the joints.[34]

Collagen glycosylation is also thought to play a major role in modulating the crosslinking process. Covalent intermolecular crosslinking of collagen is initiated after secretion in the extracellular space by oxidative deamination of key Lys and/or Hyl residues in the N- and C-telopeptide regions by lysyl oxidase (LOX) and LOX-like proteins. Then, an aldimine (Schiff base)-type crosslink can be formed spontaneously between the newly formed aldehyde in the telopeptides and the ε-amino group of the juxtaposed helical Lys or Hyl residues on the neighbouring molecule. In skeletal tissues, these divalent immature crosslinks further mature into trivalent crosslinks with pyrrole or pyridinoline structures. Crosslinking is the key molecular process to connect and stabilize the fibrils and likely to regulate fibrillogenesis and matrix mineralization (for a comprehensive review see reference [35] and references therein). Interestingly, one of the major glycosylation sites found in almost all fibrillar collagens, α1/2-Lys 87, is also one of the major intermolecular crosslinking sites. However, the role of glycosylation in crosslinking has not been demonstrated until recently. By employing a short hairpin RNA technology, we have recently showed that an altered glycosylation pattern at residue 87 has profound effects on the number and maturation of collagen crosslinks, which in turn affects the diameter of the fibrils and the mineralization process.[36,37] All these studies reinforce the critical role of collagen glycosylation in collagen functions and warrant thorough molecular characterization of this modification.

Collagen Glycosylation Analysis by Mass Spectrometry: Review of the Literature up to 2010

In early studies, mass spectrometry (MS) was employed in combination with Edman degradation and amino acid analysis for sequence determination and collagen glycosylation analysis. Typical sample preparation steps involved a series of chemical and enzymatic degradation (e.g. CNBr, collagenase, Lys C, trypsin), followed by extensive purification by molecular-sieve, reversed-phase, and ion-exchange chromatography. The collected peptides were analyzed by matrix-assisted laser desorption–ionization time-of-flight (MALDI-TOF) MS and Edman degradation.[38,39] In 1996, Mann and coworkers employed these methodologies to determine the primary structure of hydrothermal-vent worm collagen. The main difference between mammalian collagens and collagen analyzed in their study is that a large number of Hyl residues in the X–Hyl–Gly repeat are replaced by Thr. Glycosylation was found at Thr residues by MALDI-TOF and by the lack of signal in that position by Edman analysis. Similarly, Bos et al. determined the full-length sequence of porcine type X collagen. In both of these studies, glycosylated residues were inferred from both sequencing

and mass spectrometric data of individual peptides. Following protein sequencing, glycosylated peptides were characterized based on the mass difference between experimental and predicted values. Site-specific assignment of glycosylation was not possible from these data other than indirect observation by Edman degradation analysis, i.e. absence of the corresponding non-glycosylated residues.[38,39]

In another study, it was shown that interaction of CD44 on melanoma cells with the α1(IV)1263–1277 region of basement membrane collagen is modulated by glycosylation of the peptide. In this study, glycosylated peptides corresponding to those derived from type IV collagen were synthesized by fluorenylmethyloxycarbonyl (Fmoc) chemistry and characterized by MALDI-TOF MS. As matrices, sinapinic acid or 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxy-benzoic acid were used.[33]

Van den Steen et al. analyzed the degradation products of denatured type II collagen by gelatinase B[34,40] to identify possible collagen epitopes involved in pathogenesis of autoimmune rheumatoid arthritis. In the initial report,[40] peptides resulting from treatment of type II collagen with gelatinase B were separated by off-line reversed-phase high-performance liquid chromatography (RP-HPLC), and individual fractions were analyzed by Edman degradation sequencing and electro-spray ionization (ESI)–ion trap MS. Amino acid modifications including proline (Pro) and lysine (Lys) hydroxylation and glycosylated Hyl were determined based on both MS data and N-terminal amino acid sequence by Edman degradation. The presence of hexose was determined in the MS-MS mode; however, glycosylation-site determination was based on Edman sequencing, where the chromatographic retention time of the modified amino acid was shifted compared with that of the non-modified residue. This is, again, not a direct determination of such modifications. In a follow-up study published in 2004,[34] the same type II collagen digested by gelatinase B was analyzed by HPLC with post-column flow splitting; the eluent was diverted either to a fraction collector, or to a Bruker Esquire ion-trap mass spectrometer. The gelatinase B cleavage products were identified by their masses and by Edman degradation of collected fractions. The authors showed that the immunodominant epitope of human type II collagen is contained between residues 273–356 and is modified by partial Lys hydroxylation and Hyl glycosylation at residues 283 and 289.

In a more recent study published in 2007,[41] Henkel and Dreisewerd used cyanogen bromide-cleaved peptides of types I, III, and V collagen from embryonic calf skin to characterize non-crosslinked and crosslinked glycosylated peptides by MALDI-TOF MS. Peptide identification was performed by comparing the experimental masses with available collagen sequence data and those deduced from cDNA sequence. Several advantages were described for the use of CNBr in collagen analysis, such as selective and quantitative cleavage at Met residues and a low number of Met residues per collagen chain, which results in a limited number of cleavage products per collagen molecule. Furthermore, the masses of specific CNBr-cleaved peptides can be used as signatures for identifying collagen types, because of their unique amino acid composition.[42–44] Using this approach, the authors determined that the α1 and α2 chains of type V collagen are differentially glycosylated, with the α1(V) chain containing a higher ratio of disaccharide to monosaccharide compared with α2(V). For type I collagen, the glycosylated CNBr-peptide containing residue 87 was

detected in both Gal and GlcGal forms. Several glycosylated peptides from type III collagen were similarly identified. Although this method provides valuable information about the number of carbohydrates present on a specific CNBr-cleaved peptide, no site-specific information is available.[41] In addition, the large size of CNBr non-crosslinked and crosslinked peptides, the heterogeneity of modifications, and the low resolution of the linear TOF mass analyser make the data interpretation difficult.

Liquid Chromatography–Tandem Mass Spectrometry (LC-MS-MS)

Collagen glycosylation site mapping was advanced significantly by the use of liquid chromatography–tandem mass spectrometry (LC-MS-MS).[37,45–48] A chemo-enzymatic approach to identify glycosylated structures in collagen was recently described by Taga et al.[46] The described hydrazide-based methodology involves treatment of collagen *O*-glycopeptides with Sepharose 4B-immobilized galactose oxidase, followed by reaction of the newly introduced aldehyde functionality with the hydrazide resin. After removal of the non-glycosylated peptides, the hydrazone bond is hydrolyzed with 0.1% formic acid, the glycosylated peptides are eluted, reduced with sodium borohydride to their original form (aldehyde → alcohol conversion), and analyzed by reversed-phase LC-MS-MS. As proof-of-principle, standards of Gal–Hyl and GlcGal–Hyl were subjected to the hydrazide method and the products were identified on zwitterionic hydrophilic interaction liquid chromatography (ZICHILIC)-ESIMS–MS in the multiple reaction monitoring (MRM) mode. Type I and II collagens purified from bovine skin were digested with trypsin, and the mixtures were subjected to the ‘hydrazide method’. The eluted glycopeptides were separated on an Agilent 1200 series HPLC system and analyzed on-line with a QTRAP mass spectrometer. Using data dependent MS-MS of the two most abundant ions, the authors identified a considerable number of glycosylation sites, including some sites that have not been reported previously.[46] This is the first ‘enrichment-type’ approach described for characterization of collagen glycosylation. Although it holds great promise for identification of glycosylated structures in collagens, the main limitation of this method is the lack of ability to obtain information about the site-specific occupancy, i.e. distribution of Lys, Hyl, Gal–Hyl and GlcGal–Hyl at a particular site. This is due to the fact that non-glycosylated peptides containing the modification site in the form of Lys and Hyl cannot be oxidized by galactosyl oxidase, and hence will not bind to the hydrazide resin.

The use of MS-MS for characterization of collagen glycopeptides has great advantage over the conventional peptide mapping–Edman sequencing approach. The analysis is less time-consuming and site-specific information about various modifications can be readily obtained. In particular, for mucine *O*-glycopeptides, the characterization of *O*-glycosylation sites was drastically improved by the introduction of electron transfer dissociation (ETD).[49,50] Furthermore, on high-resolution instruments, it is possible to discriminate between residues with the same nominal mass, e.g. Leu/Ile (residue mass 113.0796 Da) and hydroxyproline (residue mass 113.0477 Da), based on their mass defects. Hence, their assignment can be made ‘*de novo*’ rather than using the cDNA-deduced protein sequence to map these modifications. This was recently shown by Yang et al. who used a Waters nanoACQUITY ultra-performance liquid chromatography (UPLC) system coupled to an LTQ-Orbitrap Velos mass spectrometer equipped with ETD, to map the Hyl glycosylation

and Pro hydroxylation sites in human recombinant- and bovine placenta-derived type V collagen. In addition to 106 Hyp residues within the –Y–Gly– consensus motif (most likely 4-Hyp), the authors identified several additional Hyp residues in the position X of the X–Y–Gly repeats. The identity of Hyp positional isomers (3-Hyp versus 4-Hyp) in the position X of X–Y–Gly could not be established from these data. The unusual presence of Hyp at these locations remains unexplained. A total of 34 GlcGal–Hyl residues were found in this study.[47]

We recently showed that collagen tryptic glycopeptides characteristically behave differently in collision-induced dissociation (CID) compared with typical *N*- and *O*-linked glycopeptides, in that the peptide backbone undergoes extensive fragmentation in competition with glycosidic bond dissociation.[48] Consequently, a significant number of *b/y* fragments are observed with similar or higher abundances than those of fragments arising from glycosidic bond cleavages. Because of the large number of Pro and Hyp residues in collagen glycopeptides, cleavages N-terminal to these sites are preferentially observed. We found that the main advantage of these competing fragmentation pathways is that peptide sequence determination and glycosylation site localization can be readily obtained from CID data, without the need of additional fragmentation techniques such as ETD. One representative CID spectrum is shown in Fig. 2, and several others can be found in reference [48]. In this example, the main dissociation pathway corresponds to cleavage of the amide bonds, giving rise to a series of glycosylated y fragments, including y_7 (*m/z* 876.51), y_{17} (*m/z* 1751.95), y_{20} (*m/z* 1977.08), y_{21} (*m/z* 2034.14), y_{23} (*m/z* 2202.23), y_{25} (*m/z* 2372.32), y_{26} (*m/z* 2469.34), and y_{27} (*m/z* 2526.35). Cleavage of the glycosidic bond is observed with lower relative abundance. We found this competing behaviour of the two dissociation channels (amide versus glycosidic bond) to be dependent on the number of ionizing protons relative to the number of basic sites, including the N-terminus, Arg, Lys, and Hyl.[48] When the number of ionizing protons is greater than the total number of Arg, Hyl, Lys, and N-termini, the amide bond is preferentially cleaved, while fragment ions containing the glycosylation site retain the glycan. In those instances when the number of ionizing protons is less than the total number of Arg, Hyl, Lys, and N-termini, collagen glycopeptides have a behaviour comparable with that of mucine *O*-glycopeptides, i.e. glycosidic bond cleavages are almost exclusively observed, and sequence assignment requires additional MS-MS experiments. Particularly interesting examples are those where the number of ionizing protons equals the number of sites (as in the example in Fig. 2). In these instances, the amide and glycosidic bonds either compete or one dissociation channel dominates the other, suggesting that additional factors may play a role. Using quantum mechanics calculations, we showed that cleavage N-terminal to Hyp requires less energy compared with Pro, whereas the glycosidic bond dissociation in Hyl *O*-glycosides requires at least the amount of energy necessary to dissociate the amide bonds N-terminal to Hyp and Pro.[48]

Semiquantitative Analysis of Collagen Glycosylation by LC-MS-MS

Recently, we characterized LH3 as the major glucosyl transferase in the type I collagen in an osteoblastic cell culture system.[36,37] The glycosylation sites in type I collagen isolated from wild-type mouse osteoblasts were identified by LC-MSMS on a nanoACQUITY UPLC–QTOF Premier mass spectrometer from tryptic digests of α1 and α2 chains, and

were assigned to the following five residues: α1– and α2–87, α1– and α2–174, and α2–219. Amino acid analysis of the total levels of Gal–Hyl and GlcGal–Hyl indicated that GlcGal–Hyl is diminished when LH3 expression is suppressed.[36] To determine the site-specific changes associated with suppression of LH3, we adopted a semiquantitative LC-MS-based approach, outlined in Fig. 3. The first step involves separation of the collagen α chains by sodium dodecyl sulfate polyacrylamide gel electrophoresis, followed by in-gel digestion of the individual α chains with trypsin, chosen because of its high specificity. Following qualitative characterization of glycosylated residues by MS-MS, the site occupancy was determined by taking into account all modifications of a specific site, including Lys, Hyl, Gal–Hyl, and GlcGal–Hyl. This was achieved by generating extracted ion chromatograms (EIC) corresponding to each modified protonated peptide species from the total ion current and integration of the corresponding signals. We found that trypsin cleavage is entirely abolished when Gal– and GlcGal– are present on Hyl. However, when a site is present as Lys or Hyl, the rate of trypsin cleavage may vary considerably. Consequently, fully cleaved peptides ending in Lys or Hyl and their miscleaved variants can be formed. The total ion current was inspected for the presence of all these species. All charge states in which a glycopeptide or non-glycosylated peptide was observed were integrated for the site-specific analysis. The major glycosylation site in bone type I collagen was found to be α1–87. This is also one of the major intermolecular helical crosslinking sites. This site is found >90% as GlcGal–Hyl in wild-type type I collagen in osteoblasts. In the LH3-suppressed clones, the relative amount of this glycoform at residue α1–87 is dramatically diminished, as shown in Fig. 4. A decrease in the GlcGal–Hyl glycoform was observed at the other glycosylation sites as well, although not as drastically as for residue α1–87. It is noteworthy that the homologous residue on the α2 chain, α2–87, is only minimally glycosylated, with ~90 % present as Hyl in these mouse-derived osteoblastic cells.

This is likely due to the specific sequence in the vicinity of the Hyl residue in mice.[36] In other species, α2–87 is also highly glycosylated (I. Perdivara, M. Yamauchi, M. Terajima, K. B. Tomer, unpubl. data). The change in glycosylation at α1–87 has significant impact on collagen fibril formation. The number and maturation of collagen crosslinks was also found to be reduced in the collagen synthesized by LH3-suppressed clones. This in turn inversely affects the diameter of the collagen fibrils and delays the mineralization process.[36,37]

One difficulty in quantitative collagen glycosylation analyses is (glyco)peptide heterogeneity arising from various degrees of proline hydroxylation. Pro residues in the sequence Pro–Gly are expected to be quantitatively 4-hydroxylated; however, major variations in the degree of hydroxylation of specific residues are found in different tissues. The functional consequences of these variations are not clear at present. One example of Pro underhydroxylation in type I collagen peptides from bovine skin is shown in Fig. 5. Peptides α1 [145–174] contain residues 168 and 174 in the form of both Pro and Hyp (168) and Lys/Hyl (174), while two Hyp residues are invariantly present at residues 156 and 165. Peptide [145–174] containing three Hyp residues and Hyl is observed as the ion of *m/z* 1282.65 (2+), and the variant containing two Hyp, Pro 168, and Lys 174 is observed as the ion of *m/z* 1266.65 (2+). Based on their chromatographic retention times, the isobaric species of *m/z* 1274.64 (2+) are differentiated into two populations, one containing Hyp

168/Lys 174 (38.40 min), and the later-eluting population containing Pro 168/Hyl 174 (39.97 min). The MS-MS spectra of the 3+ isobaric ions (*m/z* 850.09) acquired in the data-dependent mode (Fig. 5) clearly identify the location of hydroxylation, based on the fragment ions y_3 (*m/z* 275.20 –Lys 174 and 291.18 –Hyl 174) and y_7 (*m/z* 615.38). In Fig. 5, the abundances of the ions corresponding to differentially hydroxylated peptides were normalized to that of the most abundant species of *m/z* 1266.65. Moreover, residue 174 was found in the form of Hyl and Gal–Hyl in the miscleaved tryptic peptides [145–183]. Heterogeneity of these species due to partial hydroxylation at Pro 168 was observed as well (data not shown). This example illustrates the complexity of the sample and data for characterization of site-specific glycosylation occupancy in collagen. Additional details on quantitative determination of site-specific occupancy in collagen glycopeptides are presented in the supplemental information of reference [37].

Another remarkable feature of collagens is that the degree of Hyl glycosylation varies considerably depending on tissue, organ,[51–53] different locations within the same tissue, and pathophysiologic state.[27,54] It is also noteworthy that type I collagen-rich connective tissues such as skin, bone, and tendon have different physiologies and require distinct structural and mechanical features. This raises the question of the potential involvement of glycosylation in modulating collagen structure, stability, recruitment of other molecules (proteoglycans, minerals), and the final form and mechanical properties of tissues. As part of a larger study, we analyzed type I collagen from various species and tissues, including bovine bone and skin, rat-tail tendon, and mouse bone and skin. Preliminary results showing the differential glycosylation of the crosslinking residue α1–87 (Fig. 6) were obtained using the same quantitative approach described previously.

Remarkable differences were found in the occupancy of residues α1–87 between different tissues and species. Type I collagen from rat-tail tendon is the least glycosylated at α1–87, with >90% of this site present as Hyl. In contrast, bone, and skin type I collagen are mainly found hydroxylated and glycosylated at α1–87, with different ratios of GlcGal–Hyl to Gal–Hyl, depending on species and tissue (Fig. 6). Collagen from rat-tail tendon has some of the tightest packed fibrils, which indirectly suggests the abundance of complex, multivalent crosslinks such as dehydrohistidinohydroxymerodesmosine. Because α1–87 is one of the main crosslinking helical residues, the occupancy of this site might regulate the number and maturation of cross-linking events and subsequent fibril growth. This example demonstrates the potential of quantitative collagen glycosylation analysis in correlating molecular structures with macroscopic tissue properties.

One key requirement for accurate (near to biologic) site-specific quantitation by LC-MS is a reasonable signal-to-noise (S/N) ratio and good peak shape for the peptide species containing the modifications of interest. This analysis also requires raw (non-enriched) digests, so that sample loss is minimal. Because integration is based on generating post-data-acquisition EIC, it is still puzzling how differences in ionization efficiencies of glycosylated/non-glycosylated and long/short glycopeptides may affect these analyses. Nevertheless, achieving good chromatographic resolution of the collagen tryptic digests is crucial to ensure that glycosylated species with various degrees of heterogeneity, as described previously, are detected and accounted for.

Hence, we sought to explore the potential of nanoscale two dimensional reversed phase–reversed phase (RP-RP) separations at high/low pH in the characterization of collagen site-specific glycosylation. This technology has recently become commercially available (Waters Corporation), and was originally developed for proteomics studies.[55,56] Orthogonal peptide separations are carried out in 20 mM ammonium formate in the mobile phase (pH 10, first dimension), followed by a 10-fold online dilution and conventional RP separation at pH 2–3 in 0.1 % formic acid (second dimension). RP-RP experiments consist of alternating high pH and low pH gradients sent from two different binary pump systems on two different columns, online connected through a trapping column. For example, a three-fraction experiment at high pH consists of three distinct ramps of the organic solvent from 1 to 13.1 %, from 1 to 17.7 %, and from 1 to 65 % respectively, followed by isocratic elution at 13.1, 17.7, and 65 % organic respectively, and return to the initial 1 % organic solvent. The eluted fraction after one high-pH gradient is then 10-fold diluted in the initial solvent of the lowpH dimension, trapped, and then directed into the analytical column where the conventional low-pH separation is performed. Preliminary studies of collagens using nanoscale RP-RP separations demonstrated great potential of this technology for site-specific collagen glycosylation analyses. Two examples are shown in Fig. 7. In Fig. 7a, EIC obtained for three (glyco) peptide ions from bovine skin collagen using one dimensional (1D-nanoUPLC, left two columns) and two dimensional RP-RP (2D-nanoUPLC, right two columns) are compared. For 1D-nanoUPLC, two different gradients were tested, a 90- and a 120-min gradient. For 2D-nanoUPLC, two experiments with three and five fractions in the high-pH dimension were performed. In each column, the ion abundances of the two lower peptides α2 [928–963] containing residue 933 as Gal–Hyl and Hyl respectively were normalized to the abundances of the glycopeptide α2 [76–90] (GlcGal–Hyl at α2–87). This species was observed with good S/N in both 1D and 2D experiments. In the 1D experiments, peptides α2 [928–963] were observed at noise level, and were characterized by broad chromatographic peaks. In the 2D experiment with three fractions in the first dimension, a marked improvement in the S/N of both Hyl and Gal–Hyl-containing peptides was observed, whereas all species, i.e. peptides α2 [76–90] and α2 [928–963], eluted in fraction no. 3. When five fractions were set up in the first dimension, the S/N of peptides α2 [928–963] improved even more, whereas all analyzed species were found in fraction no. 5 of the high-pH experiment. Based on these data, site α2–933 (also known to participate in intermolecular crosslinking) appears to be largely present as Hyl with minute amounts of Gal–Hyl. The presence of Lys at this site was not yet interrogated, whereas GlcGal–Hyl was present at very low abundances.

In the preparation of type I collagen from tissues, other collagen types such as types III and V may often be co-purified, although they are normally present at much lower levels than type I. The example shown in Fig. 7b demonstrates that the S/N for glycopeptide ions α1 [76–90] from type III collagen containing the homologous crosslinking residue 87 in the form of Gal–Hyl and GlcGal–Hyl is significantly improved in the 2D compared with the 1D LC experiments. Similarly to Fig. 7a, the abundances of these species were normalized to that of glycopeptide α2 [76–90] from type I, invariably detected with good S/N in all experiments. Within RP-RP separations, the experiment with five fractions in the first dimension was qualitatively superior to the one with three fractions, because of a >2-fold

improvement in the S/N for the glycopeptide ions from type III collagen in the former compared with the latter (see Fig. 7b). Furthermore, in spite of sequence similarities, glycopeptides α2(I) [76–90] eluting in fraction no. 5 (at 65 % acetonitrile) have different retention characteristics under high-pH conditions compared with glycopeptides α1(III) [76–90] eluting in fraction no. 4 (at 20.4 % acetonitrile). These analyses demonstrate the usefulness of RP-RP separations in the detection of peptides with low levels of modifications, without additional fractionation or enrichment. This is due in part to the increased peak capacity resulting from the use of a second chromatographic dimension, which in turn might resolve some ion-suppression effects of the low-abundance species. This technology is currently being explored for quantitative site-specific glycosylation studies.

Conclusions

This review recapitulates the recent advances in collagen glycosylation analysis by mass spectrometry. In the literature before 2010, glycosylation was determined mainly from low-resolution mass spectrometric data in combination with Edman analysis. Those analyses were time-consuming and the information was obtained rather indirectly. The use of MS-MS (CID and ETD) and the availability of high-resolution instruments greatly expanded the potential to directly characterize the glycosylation pattern. The complexity of collagen samples is related to the heterogeneity of its non-stoichiometric post-translational modifications. Collagen site-specific glycosylation analysis by LC-MS provides unique molecular information that can be correlated with tissue- and species-specific information. Furthermore, nanoscale RP-RP separations at high/low pH with online dilution have great potential to improve the detection of collagen glycopeptides containing low levels of glycosylation. One of the biggest challenges is to obtain ‘absolute’ quantitative glycosylation data, e.g. the number of glycosylated residues per collagen molecule. With the availability of a large variety of commercial platforms for LC-MS-MS, we foresee an expansion of the research in the area of collagen glycosylation, in particular in the context of collagen-related diseases.

Acknowledgments

The authors would like to acknowledge the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Project no. ES 050171 to KBT) and the NIH funding grant AR060978 (MY). We would like to thank Dr Nancy Pleshko (Temple University, PA) for providing collagen samples, Drs Masahiko Terajima (University of North Carolina, NC) and Marnisa Sricholpech (Srinakarinwirot University, Bangkok, Thailand) for isolation and purification of collagen, and the NIEHS Protein Microcharacterization Facility for the assistance with the in-gel digests.

Biography

Irina Perdivara received her Diploma in Chemistry (2006) and Ph.D. with Honors in Analytical Biochemistry (2009) from the University of Konstanz, Germany. Her research at NIEHS is focussed on structure determination of proteins involved in aging, autoimmune, neurodegenerative and bone diseases by high-performance liquid chromatography – tandem mass spectrometry. Her major area of interest is characterization of protein glycosylation at the glycopeptide level in raw proteolytic mixtures to understand its functional implications in health and disease. Dr Perdivara has set the foundation for quantitative site-specific collagen glycosylation analysis, which has led to significant progress in the research areas of bone physiology and bone-related diseases.

Mitsuo Yamauchi received a DDS (1976) and a Ph.D. in Biochemistry (1983) from the Tokyo Medical and Dental University, Tokyo, Japan. During his residency in Oral Surgery at the same university, he started research on collagen and became a post-doctoral fellow in the Dental Research Center at the University of North Carolina (UNC) at Chapel Hill under the late Professor Gerald L. Mechanic (1978–81). He is currently Sunstar Distinguished Professor of Oral Biology and Professor of Periodontology at UNC at Chapel Hill. His research interests include the biology and chemistry of collagen post-translational modifications, biological functions of matrix proteins in mineralization and aging, and bone-tissue engineering.

Dr. Tomer’s research has focussed on characterization of biomolecules involved in immunological and inflammatory responses to environmental exposure. In over 300 research

articles, he has reported the development and application of new techniques – nanoflow capillary LC, direct analysis of affinity-bound analytes by MALDI-MS, on-line microdialysis-MS-MS and on-line capillary electrophoresis-MS – and major applications – protein sequences, PTMs, protein structures, epitope mapping, and small-molecule quantitation. Dr Tomer retired at the end of 2011, but remains actively engaged in mass spectrometry at NIEHS, and in the continued mentoring of Fellows as they transition to the next stage in their careers.

References

[1]. Carter EM, Raggio CL. Curr. Opin. Pediatr. 2009; 21:46. doi:10.1097/MOP.0B013E32832185C5. [PubMed: 19253462]

[2]. Gordon MK, Hahn RA. Cell Tissue Res. 2010; 339:247. doi:10.1007/S00441-009-0844-4. [PubMed: 19693541]

[3]. Leitinger B. Ann. Rev. Cell Develop. Biol. 2011; 27:265. doi:10.1146/ANNUREV-CELLBIO-092910-154013.

[4]. Kivirikko KI, Myllyla R. Methods Enzymol. 1982; 82:245. doi:10.1016/0076-6879(82)82067-3. [PubMed: 6210830]

[5]. Kivirikko KI, Myllyla R. Methods Enzymol. 1987; 144:96. doi:10.1016/0076-6879(87)44175-X. [PubMed: 3041180]

[6]. Kolble K, Reid KB. Behring Inst. Mitt. 1993; 93:81. [PubMed: 8172588]

[7]. Grassmann W, Schleich H. Biochem. Z. 1935; 277:320.

[8]. Butler WT, Cunningham LW. J. Biol. Chem. 1966; 241:3882. [PubMed: 4288358]

[9]. Spiro RG. J. Biol. Chem. 1967; 242:4813. [PubMed: 6070267]

[10]. Liefhebber JMP, Punt S, Spaan WJM, van Leeuwen HC. BMC Cell Biol. 2010; 11:33. [PubMed: 20470363]

[11]. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T. Mol. Cell. Biol. 2009; 29:943. doi: 10.1128/MCB.02085-07. [PubMed: 19075007]

[12]. Myllylä R, Wang CG, Heikkinen J, Juffer A, Lampela O, Risteli M, Ruotsalainen H, Salo A, Sipilä L. J. Cell. Physiol. 2007; 212:323. doi:10.1002/JCP.21036. [PubMed: 17516569]

[13]. Wang CG, Luosujarvi H, Heikkinen J, Risteli M, Uitto L, Myllyla R. Matrix Biol. 2002; 21:559. doi:10.1016/S0945-053X(02)00071-9. [PubMed: 12475640]

[14]. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, Myllylä R. J. Biol. Chem. 2000; 275:36158. doi:10.1074/JBC.M006203200. [PubMed: 10934207]

[15]. Mizuno K, Adachi E, Imamura Y, Katsumata O, Hayashi T. Micron. 2001; 32:317. doi:10.1016/S0968-4328(00)00036-6. [PubMed: 11006511]

[16]. Shinkai H, Yonemasu K. Biochem. J. 1979; 177:847. [PubMed: 444207]

[17]. Moro L, Romanello M, Favia A, Lamanna MP, Lozupone E. Calcif. Tissue Int. 2000; 66:151. doi:10.1007/S002230010030. [PubMed: 10652964]

[18]. Toole BP, Kang AH, Trelstad RL, Gross J. Biochem. J. 1972; 127:715. [PubMed: 4651137]

[19]. Risteli M, Ruotsalainen H, Salo AM, Sormunen R, Sipila L, Baker NL, Lamandé SR, Vimpari-Kauppinen L, Myllylä R. J. Biol. Chem. 2009; 284:28204. doi:10.1074/JBC.M109.038190. [PubMed: 19696018]

[20]. Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, Robins SP, Myllylä R. Am. J. Hum. Genet. 2008; 83:495. doi:10.1016/J.AJHG.2008.09.004. [PubMed: 18834968]

[21]. Bateman JF, Mascara T, Chan D, Cole WG. Biochem. J. 1984; 217:103. [PubMed: 6421277]

[22]. Brenner RE, Vetter U, Nerlich A, Worsdorfer O, Teller WM, Muller PK. Eur. J. Clin. Invest. 1990; 20:8. doi:10.1111/J.1365-2362.1990.TB01785.X. [PubMed: 2108039]

[23]. Brinckmann J, Notbohm H, Tronnier M, Acil Y, Fietzek PP, Schmeller W, Müller PK, Bätge B. J. Invest. Dermatol. 1999; 113:617. doi:10.1046/J.1523-1747.1999.00735.X. [PubMed: 10504450]

[24]. Cetta G, De Luca G, Tenni R, Zanaboni G, Lenzi L, Castellani AA. Connect. Tissue Res. 1983; 11:103. doi:10.3109/03008208309004847. [PubMed: 6224635]

[25]. Lehmann HW, Wolf E, Roser K, Bodo M, Delling G, Muller PK. J. Cancer Res. Clin. Oncol. 1995; 121:413. doi:10.1007/BF01212948. [PubMed: 7635871]

[26]. Michalsky M, Norris-Suarez K, Bettica P, Pecile A, Moro L. Biochem. Biophys. Res. Commun. 1993; 192:1281. doi:10.1006/BBRC.1993.1555. [PubMed: 8507198]

[27]. Tenni R, Valli M, Rossi A, Cetta G. Am. J. Med. Genet. 1993; 45:252. doi:10.1002/AJMG.1320450219. [PubMed: 8456811]

[28]. Yamauchi M, Sricholpech M. Essays Biochem. 2012; 52:113. [PubMed: 22708567]

[29]. Bhadriraju K, Chung KH, Spurlin TA, Haynes RJ, Elliott JT, Plant AL. Biomaterials. 2009; 30:6687. doi:10.1016/J.BIOMATER IALS.2009.08.036. [PubMed: 19762078]

[30]. Jurgensen HJ, Madsen DH, Ingvarsen S, Melander MC, Gardsvoll H, Patthy L, Engelholm LH, Behrendt N. J. Biol. Chem. 2011; 286:32736. doi:10.1074/JBC.M111.266692. [PubMed: 21768090]

[31]. Vogel W, Gish GD, Alves F, Pawson T. Mol. Cell. 1997; 1:13. doi:10.1016/S1097-2765(00)80003-9. [PubMed: 9659899]

[32]. Xu HF, Raynal N, Stathopoulos S, Myllyharju J, Farndale RW, Leitinger B. Matrix Biol. 2011; 30:16. doi:10.1016/J.MATBIO.2010.10.004. [PubMed: 21044884]

[33]. Lauer-Fields JL, Malkar NB, Richet G, Drauz K, Fields GB. J. Biol. Chem. 2003; 278:14321. doi:10.1074/JBC.M212246200. [PubMed: 12574156]

[34]. Van den Steen PE, Proost P, Brand DD, Kang AH, Van Damme J, Opdenakker G. Biochemistry. 2004; 43:10809. doi:10.1021/BI0493665. [PubMed: 15311942]

[35]. Yamauchi M, Shiiba M. Methods Mol. Biol. 2008; 446:95. doi:10.1007/978-1-60327-084-7_7. [PubMed: 18373252]

[36]. Sricholpech M, Perdivara I, Nagaoka H, Yokoyama M, Tomer KB, Yamauchi M. J. Biol. Chem. 2011; 286:8846. doi:10.1074/JBC.M110.178509. [PubMed: 21220425]

[37]. Sricholpech M, Perdivara I, Yokoyama M, Nagaoka H, Terajima M, Tomer KB, Yamauchi M. J. Biol. Chem. 2012; 287:22998. doi:10.1074/JBC.M112.343954. [PubMed: 22573318]

[38]. Bos KJ, Rucklidge GJ, Dunbar B, Robins SP. Matrix Biol. 1999; 18:149. doi:10.1016/S0945-053X(99)00015-3. [PubMed: 10372555]

[39]. Mann K, Mechling DE, Bachinger HP, Eckerskorn C, Gaill F, Timpl R. J. Mol. Biol. 1996; 261:255. doi:10.1006/JMBI.1996.0457. [PubMed: 8757292]

[40]. Van Den Steen PE, Proost P, Grillet B, Brand DD, Kang AH, Van Damme J, Opdenakker G. FASEB J. 2002; 16:379. doi:10.1096/FJ.01-0688COM. [PubMed: 11874987]

[41]. Henkel W, Dreisewerd K. J. Proteome Res. 2007; 6:4269. doi:10.1021/PR070318R. [PubMed: 17939700]

[42]. Rauterberg J, Kuhn K. Eur. J. Biochem. 1971; 19:398. doi:10.1111/J.1432-1033.1971.TB01329.X. [PubMed: 5102923]

[43]. Rhodes RK, Gibson KD, Miller EJ. Biochemistry. 1981; 20:3117. doi:10.1021/BI00514A020. [PubMed: 7248273]

[44]. Rhodes RK, Miller EJ. J. Biol. Chem. 1979; 254:12084. [PubMed: 500696]

[45]. Zhang G, Sun A, Li W, Liu T, Su Z. J. Chromatogr. A. 2006; 1114:274. doi:10.1016/J.CHROMA.2006.03.039. [PubMed: 16600269]

[46]. Taga Y, Kusubata M, Ogawa-Goto K, Hattori S. Mol. Cell. Proteomics. 2012; 11 M111.010397. doi:10.1074/MCP.M111.010397.

[47]. Yang CX, Park AC, Davis NA, Russell JD, Kim B, Brand DD, Lawrence MJ, Ge Y, Westphall MS, Coon JJ, Greenspan DS. J. Biol. Chem. 2012; 287:40598. doi:10.1074/JBC.M112.406850. [PubMed: 23060441]

[48]. Perdivara I, Perera L, Sricholpech M, Terajima M, Pleshko N, Yamauchi M, Tomer KB. J. Am. Soc. Mass Spectrom. 2013 advance article. doi:10.1007/S13361-013-0624-Y.

[49]. Peter-Katalinic J. Method Enzymol. 2005; 405:139. doi:10.1016/S0076-6879(05)05007-X.

[50]. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Proc. Natl. Acad. Sci. USA. 2004; 101:9528. doi:10.1073/PNAS.0402700101. [PubMed: 15210983]

[51]. Pinnell SR, Fox R, Krane SM. Biochim. Biophys. Acta. 1971; 229:119. doi: 10.1016/0005-2795(71)90325-4. [PubMed: 5101159]

[52]. Schofiel JD, Freeman IL, Jackson DS. Biochem. J. 1971; 124:467. [PubMed: 5002672]

[53]. Segrest JP, Cunningh LW. J. Clin. Invest. 1970; 49:1497. doi:10.1172/JCI106367. [PubMed: 5431661]

[54]. Dominguez LJ, Barbagallo M, Moro L. Biochem. Biophys. Res. Commun. 2005; 330:1. doi: 10.1016/J.BBRC.2005.02.050. [PubMed: 15781223]

[55]. Gilar M, Olivova P, Daly AE, Gebler JC. J. Sep. Sci. 2005; 28:1694. doi:10.1002/JSSC. 200500116. [PubMed: 16224963]

[56]. Gilar M, Olivova P, Daly AE, Gebler JC. Anal. Chem. 2005; 77:6426. doi:10.1021/AC050923I. [PubMed: 16194109]

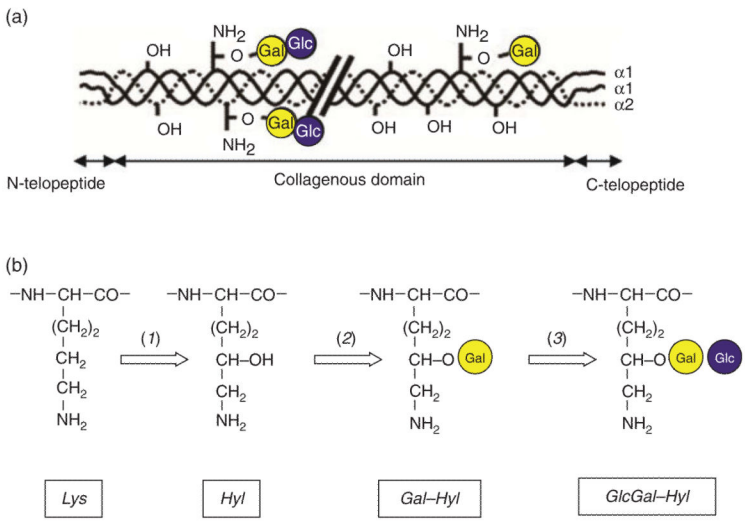

Fig. 1.

(a) Schematic representation of a type I collagen molecule formed by intertwining two α1 and one α2 chains: unstructured N-terminal telopeptide region, followed by the triple helix (‘collagenous domain’) and by the C-terminal telopeptide region. The OH groups indicate the presence of hydroxylated residues Hyp and Hyl. Glycosylation of Hyl is indicated with light grey (yellow) (Gal) and dark grey (blue) circles (Glc). (b) The sequence of post-translational events involved in modification of Lys residues in collagen: (1) lysine hydroxylation; (2) addition of Gal to form Ga–Hyl, and (3) addition of Glc to form GlcGal–Hyl.

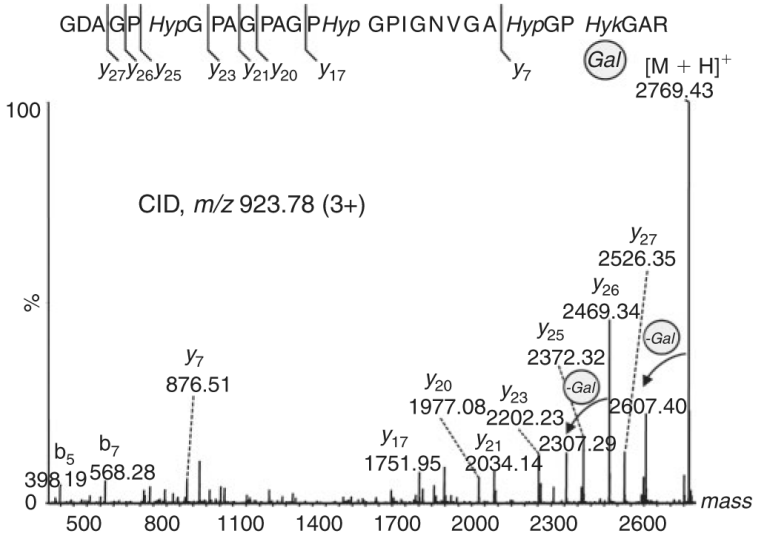

Fig. 2.

Deconvoluted collision-induced dissociation (CID) spectrum of the glycopeptide ion of *m*/*z* 923.78 (3+) observed in a bovine skin tryptic digest, corresponding to the indicated peptide modified with Gal–Hyl. The spectrum was obtained with a collision energy of 30 V.

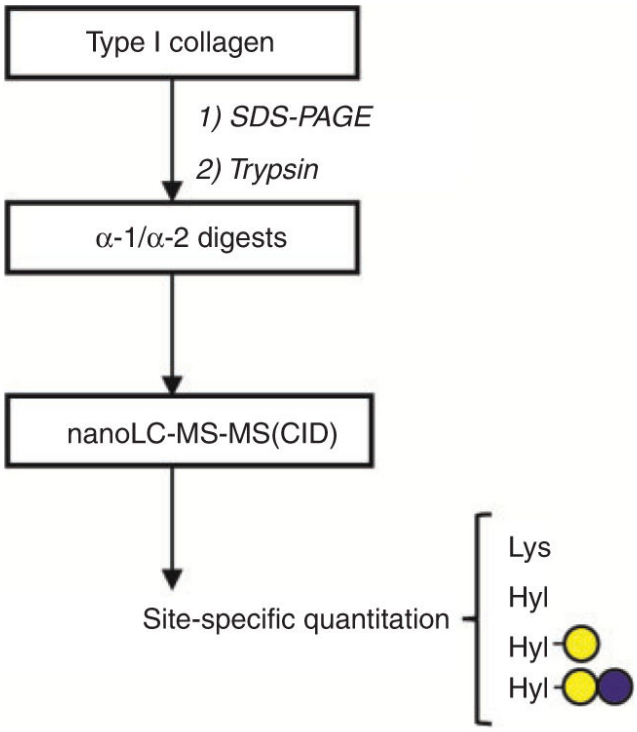

Fig. 3.

Analytical strategy employed for site-specific glycosylation analysis by liquid chromatography–tandem mass spectrometry (LC-MS-MS).

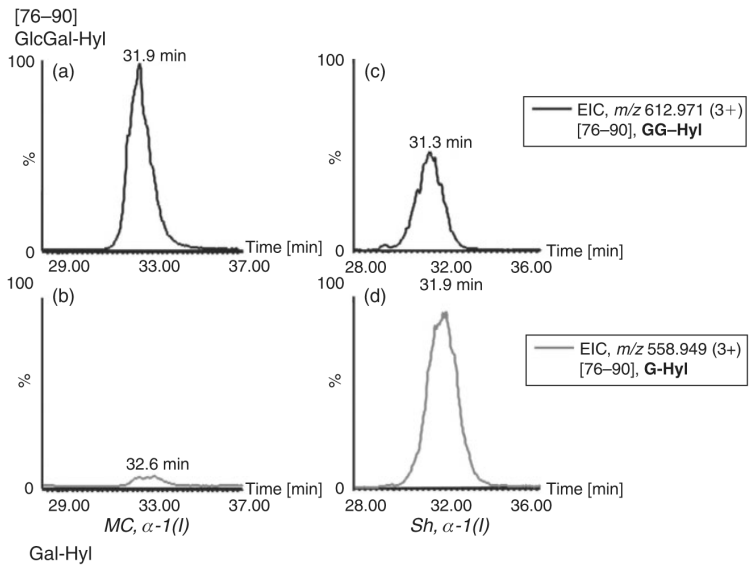

Fig. 4.

Extracted ion chromatograms (EIC) of *m*/*z* 612.97 (3+) and 558.95 (3+), corresponding to glycopeptides α-1(I) [76–90], containing residue 87 as GlcGal–Hyl and Gal–Hyl respectively: **A** and **B**, wild type (MC) collagen; **C** and **D**, collagen isolated from lysyl hydroxylase 3 (LH-3)-suppressed mouse osteoblastic cells (Sh). For each collagen sample, the EIC were normalized to the abundance of the most abundant glycoform. This figure is reproduced from Sricholpech et al. *J. Biol. Chem.* **2012**, 287(27), 22998.

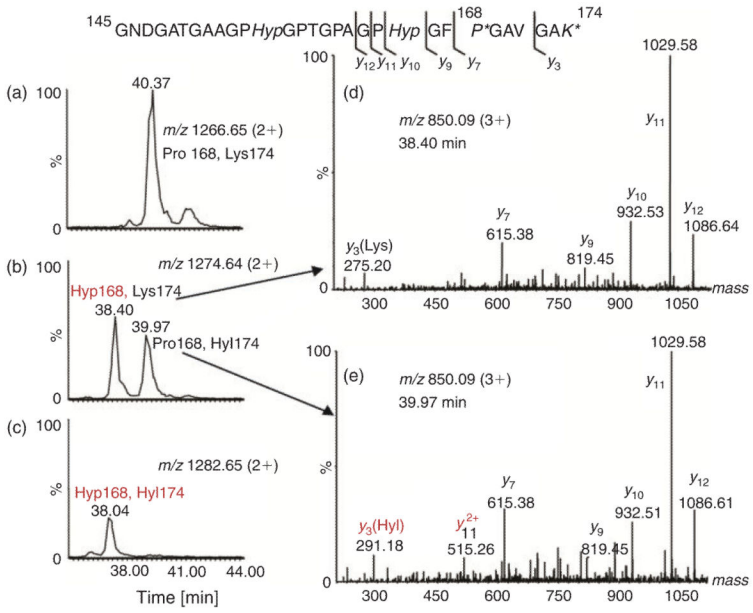

Fig. 5.

Extracted ion chromatograms (EIC) of peptide ions corresponding to bovine skin peptides α1(I) [145–174] containing different degrees of hydroxylation: (a) *m*/*z* 1266.65 (2+); (b) *m*/*z* 1274.64 (2+); and (c) *m*/*z* 1282.65 (2+). Partially hydroxylated residues are shown in red; (d) and (e) deconvoluted region (*m*/*z* 200–1100 Da) of the collision-induced dissociation (CID) spectra of the ions of *m*/*z* 850.09 (3+), corresponding to the isobaric species [145–174] containing hydroxylation at residue 168 (d) or 174 (e). The ions indicating the location of hydroxylation are highlighted in red.

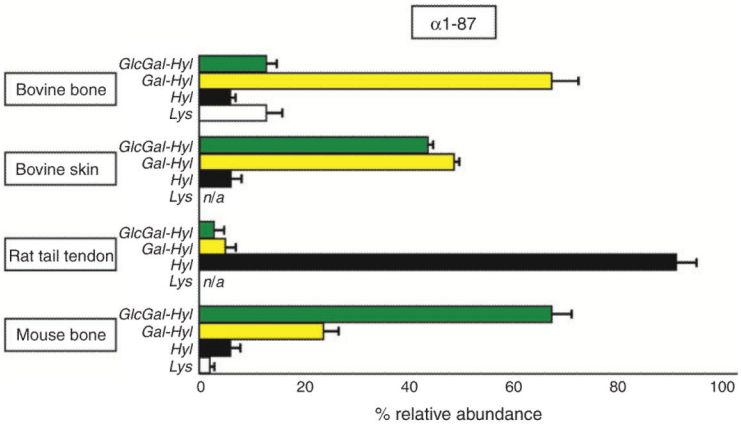

Fig. 6.

Distribution of Lys, Hyl, Gal–Hyl and GlcGal–Hyl at residue α1–87 in bovine bone, bovine skin, rat-tail tendon, and mouse bone.

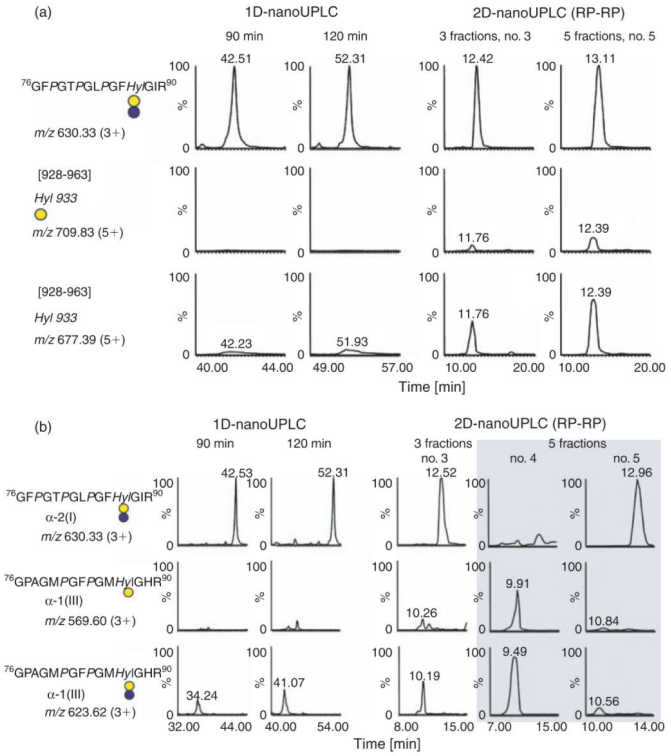

Fig. 7.

Comparison of nanoscale 1D and 2D UPLC-MS (ultra-performance liquid chromatography–mass spectrometry) for detection of (a) collagen glycopeptides containing low levels of modifications; shown here are peptides α2(I) [928–963] containing Hyl and Gal–Hyl respectively at residues 933; and (b) glycopeptides from types I and III collagen, having different ion abundances. In both (a) and (b), bovine skin glycopeptide α2(I) [76–90] modified with GlcGal–Hyl was chosen as reference. Extracted ion chromatograms (EICs) in the same column were normalized to the abundance of peptide α2(I) [76–90]. Light grey (Yellow), Gal; dark grey (blue), Glc. For more details, see text.