PLOS ONE

RESEARCH ARTICLE

Heterogeneity in proline hydroxylation of fibrillar collagens observed by mass spectrometry

Michele Kirchner[1,2], **Haiteng Deng**[3¤], **Yujia Xu**[1,2]*

1 Department of Chemistry, Hunter College of CUNY, New York, NY, United States of America, **2** The Graduate Center, The City University of New York, New York, NY, United States of America, **3** Proteomics Resource Center, The Rockefeller University, New York, NY, United States of America

¤ Current address: School of Life Sciences, Tsinghua University, Haidian District, Beijing, China
* yujia.xu@hunter.cuny.edu

Citation: Kirchner M, Deng H, Xu Y (2021) Heterogeneity in proline hydroxylation of fibrillar collagens observed by mass spectrometry. PLoS ONE 16(8): e0250544. https://doi.org/10.1371/journal.pone.0250544

Editor: Joseph Banoub, Fisheries and Oceans Canada, CANADA

Received: April 7, 2021

Accepted: June 28, 2021

Published: August 31, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The study was supported by NIH grant SC1 GM121273 and 5SC3GM083822 (Y.X.), NSF grant CHEM 1022120 (Y.X.), and the PSC-CUNY (The City University of New York) award 61826-00 39 (Y.X.). This publication was also made possible by Grant Number RR003037 from the National Center for Research Resources (NCRR), a component of the National Institute of Health (NIH).

Abstract

Collagen is the major protein in the extracellular matrix and plays vital roles in tissue development and function. Collagen is also one of the most processed proteins in its biosynthesis. The most prominent post-translational modification (PTM) of collagen is the hydroxylation of Pro residues in the Y-position of the characteristic (Gly-Xaa-Yaa) repeating amino acid sequence of a collagen triple helix. Recent studies using mass spectrometry (MS) and tandem MS sequencing (MS/MS) have revealed unexpected hydroxylation of Pro residues in the X-positions (X-Hyp). The newly identified X-Hyp residues appear to be highly heterogeneous in location and percent occupancy. In order to understand the dynamic nature of the new X-Hyps and their potential impact on applications of MS and MS/MS for collagen research, we sampled four different collagen samples using standard MS and MS/MS techniques. We found considerable variations in the degree of PTMs of the same collagen from different organisms and/or tissues. The rat tail tendon type I collagen is particularly variable in terms of both over-hydroxylation of Pro in the X-position and under-hydroxylation of Pro in the Y-position. In contrast, only a few unexpected PTMs in collagens type I and type III from human placenta were observed. Some observations are not reproducible between different sequencing efforts of the same sample, presumably due to a low population and/or the unpredictable nature of the ionization process. Additionally, despite the heterogeneous preparation and sourcing, collagen samples from commercial sources do not show elevated variations in PTMs compared to samples prepared from a single tissue and/or organism. These findings will contribute to the growing body of information regarding the PTMs of collagen by MS technology, and culminate to a more comprehensive understanding of the extent and the functional roles of the PTMs of collagen.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations

Both the single letter and the three letter abbreviations of an amino acid will be used with the following additions: Hyp or O stands for 4R-hydroxylated proline and 3Hyp stands for 3-hydroxylated proline. When needed for clarity, the lower case single letter abbreviation will be used to represent the genomic DNA sequence, and upper case ones the sequence seen in the peptides.

Introduction

The high sensitivity of mass spectrometry (MS) and tandem MS sequencing (MS/MS) has led to the identification of new post-translational modifications (PTMs) in fibrillar collagen and even helped to expand the field of collagen research to include archeology for the study of ancient species [1–10]. Fibrillar collagen is the major protein of bone, skin, cartilage, and blood vessel walls, and plays critical roles in many physiological and pathological events [11, 12]. The newly discovered PTMs that are of particular interest is the 3-hydroxyproline residues (3Hyp, or 3O) in unexpected locations, since mutations in the enzymes involved in the formation of 3Hyp have been linked to severe cases of Osteogenesis Imperfecta (the brittle bone diseases) [13, 14]. Yet, emerging from further studies of 3Hyp is an increasingly more heterogeneous pattern in terms of number, location, and the percent occupancy of this PTM [3, 9]. Such varied patterns of 3Hyp make it challenging to pin down the specific molecular interactions involving 3Hyp. At the same time, collagens have been used as biomarkers for disease detection, species identification, and investigations of the involvement of collagen in cancer metastasis, tissue remodeling, the homeostasis of the extracellular matrix, and the mineralization process of bones, to name just a few [15–18]. The unpredictable nature of prolyl-3-hydroxylation has hampered the application of MS in these and other related research that rely on a precise knowledge of the genomic sequence and the PTMs of specific segments of collagen. Both the functional research of PTMs and the applications of MS for a broad range of quantitative studies of collagen depend on a comprehensive understanding of the extent and the variations of PTMs.

Collagen is a highly processed protein during its biosynthesis [11, 12]. The triple helix domain of fibrillar collagens, which include collagens type I-III, V and XI, often contains more than 1000 amino acid residues in the uninterrupted (Gly-Xaa-Yaa) repeating amino acid sequence. While the Xaa and Yaa of the (Gly-Xaa-Yaa) triad can be any amino acid, about 10–12% of the residues at each of the X- and Y-positions are Pro [19]. The major PTM of collagen is the prolyl-4-hydroxylation of Pro residues in the Y-position [11]. This is a stable, invariant modification; nearly all Pro residues in the Y-position are hydroxylated to 4R-hydroxyproline (4Hyp). Some Lys residues in the Y-position are also hydroxylated to hydroxylysine (Hyl); these are often glycosylated and/or form covalent cross-links through an oxidation process catalyzed by lysyl oxidase during tissue maturation [20–23]. Until recently, only one Pro residue in the X-position, Pro^{986} of the α1 chain of type I and type II collagen (the α1(I) chain and the α1(II) chain, respectively) was known to be a 3R-hydroxyproline (3Hyp), in which the hydroxyl group is appended on the β-carbon (the 3-position) of the pyrrole ring of Pro instead of the γ-carbon (the 4-position) as is the case of 4Hyp [24]. More recent studies using MS and MS/MS found that several other Pro in the X-position of fibrillar collagens were also hydroxylated [8, 9]; some of the X-Hyps were later confirmed to be a 3Hyp by Edman sequencing. Differing from the $3Hyp^{986}$, which is invariant and has a nearly 100% occupancy, the newly discovered X-Hyp and 3Hyp residues are often found in a mixed population having the percentage of hydroxylated moiety ranging from 10% to 80% depending on the location, the type of collagen, the tissue, and the organism [3, 5, 8, 9].

The PTMs of collagen are essential for its secretion, self-assembly, and the immune responses of tissues [9, 20, 25–29], although much of the molecular mechanisms of their involvements remain unclear. Determining the biological functions of PTMs in fibrillar collagen is often confounded by the complex structural hierarchy of collagen fibrils [11, 12]. The collagen triple helix consists of three polypeptide chains that twist about a common axis to form a rod-shaped molecule about 300 nm in length. The triple helices further self-associate laterally in a specific manner to form collagen fibrils having a unique 67 nm axially repeating structure known as the *D*-period. Any modifications of residues in X or Y positions can potentially impact the stability of the triple helix, the molecular recognition process during fibrillogenesis, and the interactions of collagen fibrils with cell receptors and other macromolecules during tissue development and function. Impaired prolyl-4-hydroxylation is the major cause of the condition of scurvy linked to the fragility in skin, blood vessels, and dentine [20]. In this case, the tissue fragility was linked to the decreased stability of the collagen triple helix due to the lack of 4Hyp. Studies using triple helical peptides have firmly established the significant stabilizing effects of a 4Hyp in the Y-position compared to that of a Pro [30–34]. The Hyl related glycosylation and cross-links were also considered an important part of the fibril stability, and the extent of the modification increases with the advance of the developmental stages [35–37]. The understanding of the function of 3Hyp is more limited, except it is important for bone health [13, 38, 39]. Eyre and colleagues postulate the newly discovered 3Hyps are involved in fibrillogenesis because the locations of some of them are approximately a *D*-period apart [5, 7]. However, considering the low occupancy at some of the locations, it remains to be evaluated at what extent of the hydroxylation the purported interactions involving 3Hyp will have a sustained impact on the fibril assembly. A systematic MS/MS characterization of rat type I collagen found an increased occupancy of 3Hyp with the developmental stages in rat-tail tendon, but the same study also reported a relatively constant extent of hydroxylation of type I collagen in bones and in skin [3]. If the prevalent presence of 4Hyp is consistent with its structural role on the overall stability of the triple helix and the collagen fibrils, the highly diverged and sporadic presence of 3Hyp may suggest a more dynamic role for this unique collagen PTM, which may involve the hydroxylation of specific X-Pro residues at specific stages of development and/or in response to specific cues of the extracellular matrix. Similar dynamic PTMs were reported to be part of the 'epigenetic code' of histone and other proteins [40].

It is often difficult to delineate the variations of a PTM as part of the dynamic epigenetic regulation from the statistical variations of the techniques used for detection and/or for sample handling [41, 42]. The situation is particularly challenging for collagen due to the repetitive sequences and the high content of Pro residues. Fragmentation of Pro containing peptides is often inefficient due to the well characterized 'proline effect' in MS/MS which predicts a biased, sequence dependent potential to fragment N-terminal to a Pro bond during collision induced dissociation (CID) [2, 9, 43–45]. In the case of the detection of 3Hyp the high frequency of the genomic sequence of pro-gly-pro-pro moiety further complicated the precise localization of the hydroxyl group without a good series of fragmented ions. Aside from functional dynamics, there is often an innate level of variations in the PTMs of a protein between individuals and between different organisms. Without a known priori on the statistical distribution of the PTMs in a specific tissue at a specific developmental stage, even a carefully designed study can only reflect a statistical snapshot. Collagens produced by recombinant systems may appear to be a well-controlled source of more homogeneous collagens. However, the expression of a foreign gene(s) can skew the PTM processes in a host cell leading to a different PTM pattern [1].

To gain a better understanding of the average impact of the prolyl-hydroxylation in both the X and the Y positions and the reliability of the detection by standard MS techniques, we

carried out MS/MS sequencing of several samples of collagen from commercial sources and of collagen isolated from tissues. Our study revealed a more varied nature of the hydroxylation of proline residues in the type I collagen and substantial differences in the hydroxylation pattern among different collagens. The productions of commercial collagens often rely on batch collection of samples from mixed sources. This 'mixed' nature, however, can potentially make the commercial collagen a good statistical representation of the overall extent of different PTMs. Additionally, the information on PTMs will also enhance the other applications of commercial collagens as standards for analytical analyses, and as extracellular matrix substitutes in various biological and biomedical studies that frequently rely on specific interactions with residues on collagen including the PTMs. Mapping out all the 3Hyp residues with the highest sensitivity and accuracy is not the main focus of this work. Rather, we seek to understand the reliability and reproducibility of the detection of unexpected hydroxylations using the standard MS approach. Furthermore, the uncertainties of PTMs on the X position complicate the fundamental premises of MS studies of collagen that assumes Pro in X-positions are unmodified, while those in the Y-position will inevitably have a mass increase of 16 due to the addition of the hydroxyl group. The finding of this work will, thus, contribute to both the understanding of the dynamics of 3Hyp and the applications of MS in other areas of collagen research.

Materials and methods

Collagen samples

Human collagen type III, human collagen type I, and rat collagen type I were purchased from Sigma. According to Sigma, human collagen type I and type III was purified from placenta, and rat collagen type I from tail tendon. The purchased collagens were solubilized in 20 mM acetic acid, pH 3, at 4˚C and 2.4 mg/mL. Fresh rat collagen type I was prepared from a single rat tail tendon following a procedure published by Dr. Sergey Leikin's group, and the precipitated collagen was solubilized in 20 mM acetic acid, pH 3, at 4˚C and 3.0 mg/mL [46]. Collagen was mixed with 5X SDS sample loading buffer containing 60 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 350 mM DTT, and 0.1% bromophenol blue, and was run on a 4–20% Precise gel or a 7.5% SDS PAGE gel and stained with coomassie blue. Bands of alpha chains were excised from the gels and submitted to The Rockefeller University for in-gel digestion and mass spectrometry analysis.

In-gel trypsin digestion

The gel bands were reduced with dithiothreitol for 45 minutes at 55˚C, and alkylated with iodoacetamide for 30 minutes at room temperature in the dark. 10 μL of 0.02 μg/μL trypsin in 50 mM NH_4HCO_3/0.1% octyl glucopyranoside (OGP)/ 5 mM calcium chloride was used to digest each sample overnight at 37˚C in 50 mM NH_4HCO_3. The digestion was stopped with the addition of 5 μL of 10% acetic acid, and the peptides were extracted first with 30% ACN/ 5% TFA and then with 50% ACN/5% TFA. The samples were dried down in a Speed Vac to a few microliters.

Trypsin in-solution digestion

The collagen solution was heat denatured and digested overnight with trypsin in 25 mM ammonium bicarbonate at 37˚C using an enzyme to substrate ratio of 1:20 or 1:50 weight/ weight. The reaction was terminated by bringing the solution to a final concentration of 0.1% trifluoroacetic acid.

MALDI-TOF MS analysis

Matrix, α-cyano-4-hydroxycinnamic acid, was prepared as a saturated solution in 50% acetonitrile/0.1% trifluoroacetic acid. Solution digests of collagen were spotted 1:1 with matrix onto a sample plate and allowed to dry. All spectra were acquired using a Voyager-DE STR mass spectrometer (PE Biosystems, Foster City, CA) equipped with a pulse nitrogen laser (λ = 337 nm, 3 Hz frequency) in the reflectron positive ion, delayed extraction mode. Spectra from 100 individual laser shots were averaged.

LC-MS/MS analysis

The in-gel digests and solution digest samples were chromatographed using a C_{18} column on a Dionex HPLC eluted with a gradient of 0.1% formic acid and 100% ACN and introduced into a mass spectrometer. The sequencing was done at The Rockefeller University Proteomics Resource Center. All samples underwent at least two rounds of sequencing and the data presented in the following sections are the combined results of these sequencing efforts.

Data handling and database search

The .raw data files were converted to .dta or .mgf files. For protein identification, a Mascot (Matrix Science) search was performed. The databases used were the Swiss Prot database and the User 0710 (in house database at Rockefeller's Proteomics Resource Center) containing human collagen type I and type III, and rat collagen type I. Oxidized methionine, proline and lysine were designated as variable modifications.

Data analysis

The data analysis was carried out by the Proteomics Resource Center at The Rockefeller University using standard software packages. Specifically, the protein search was done using MudPIT protein scoring, and a decoy database. The reliability of the MS/MS largely depends on the detection of the fragmented ions. Peptides with a score of 40 or higher were selected and an effort was made to manually check the identified ions in the sequencing outcome. For the freshly prepared rat tail, a summary report file was created using Discoverer version: 1.3.0.339 with the precursor mass tolerance set to 20 ppm, and the fragment mass tolerance was set to 0.5 Da. The signal to noise threshold was 1.5. To aid in correct spectrum identification, the Percolator strict false discovery rate was set to 0.01 and the maximum delta correlation was set to 0.05. The validation was based on a *q* value of 0.01 and the mass precision was set to 4 ppm.

The statistical significance of the results is reflected in the scoring of the standard software packages used for analysis. A detailed evaluation of the statistical features of the scoring schemes is beyond the scope of the current work; some systematic studies of these packages can be found in the literature [47–51]. In order to gain a qualitative understanding of the reproducibility of the sequence efforts, each collagen sample was sequenced at least twice, with the sample of human type III collagen sequenced four times. The reproducible observations were indicated in the bold font in the results section; the complete sequencing results are included in the S1 File.

Results

The variations in the hydroxylation of collagen

The heterogeneity of the hydroxylation of collagen α-chains can be observed at different levels by mass spectrometry (MS). The sequencing by MS is carried out on trypsin digested peptides of collagen which usually range from 6–30 amino acid residues in length. The observed mass

of one tryptic peptide that has undergone heterogeneous hydroxylation will resemble a mixture of species with mass differences of 16 (replacing–H by–OH) or multiples of 16. The presence of such mixed species was frequently observed in MALDI MS spectra of a total collagen digest (Fig 1). The exact number of such peak-clusters varies among different collagens, and/or the different preparations of the same collagen. The observation of such peaks by MALDI MS depends on the trypsin digestion reaction and the signal level in the MALDI MS spectra. Only those peaks with the highest level of signal are labeled in Fig 1; many others are present with low signal levels. Although it is not possible to identify these peptides by MALDI MS *per se*, given the widespread existence of peaks with mass variants of 16, it is unlikely for these peaks to be caused by the coincidental mass variations of different tryptic peptides. Rather, these peaks point to a mixed population of tryptic peptides with incomplete and/or 'over' hydroxylation of amino acid residues. The MS/MS sequencing study further supports this conclusion.

The O_x and the P_y residues identified by MS/MS sequencing of the collagens

For the clarity and the convenience of data presentation we will use O_x (or X-Hyp) and P_y (or Y-Pro), respectively, for a hydroxylated Pro in the X position and an unhydroxylated Pro in the Y-position to highlight the unusual hydroxylation results; the normal symbols of P (or Pro) and O (or Hyp) are used for Pro and 4-hydroxyproline in X and Y positions, respectively. Even by MS/MS sequencing, the exact modification often cannot be resolved with certainty with the mass information alone. In case a +16 mass is observed for a fragment of pro-gly-pro or pro-pro-gly sequence, for example, it is generally necessary to assume the hydroxylation of Pro is in the Y-position and not in the X-position in order to resolve the mass variations. Such fragments are common because of the high content of Pro in collagen and the proline-effect of MS/MS-sequencing [2, 9, 43–45]. In compiling the sequencing data the 'theoretical mass' is calculated assuming *all* Pro residues in the Y-position are hydroxylated. Thus, a mass variation of -16 reflects an incomplete hydroxylation of Y-Pro residues, while that of +16 indicates an *additional* hydroxylation beyond the usual Hyp at Y-positions. In addition to Pro in an X position, a modified Lys (Hyl) in a Y-position or an oxidized Met (M_{ox}) would also cause a mass change of +16 compared to their unmodified counterpart [52]. A C-terminal Y-position Lys in a peptide can potentially be a hydroxylated Lys and contribute to the +16 mass variation. Considering the 7-fold decrease in trypsin susceptibility to Hyl relative to that of Lys, however, a C-terminal Hyl of a tryptic digest is an unlikely event [53]. The C-terminal Lys residues are, thus, usually taken as one that is not hydroxylated with one exception: Hyl^{87} of the α2 chain of rat tail type I collagen (see later sections). The hydroxylation of this Lys was supported unambiguously by the +16 data of fragmented ions: all Y-Pro residues of this tryptic peptide are all fully occupied and there is no Pro in the X-position. The Lys in the equivalent position of human collagen was found to be hydroxylated by other methods [2]. Hyl will not be seen in the sequencing results if glycosylated or cross-linked to neighboring peptides.

In the following, we only report the O_x and P_y residues that are unambiguously supported by a series of fragmented b- and/or y- ions. One typical example of unusual hydroxylation is given in Fig 2. There are 3 Pro residues in this 19-residue peptide from position 435–453 of the α_1(I) chain of collagen isolated from a single rat tail tendon (srtt): two in the X-position (p^{446}, p^{449}) and one in Y-position (p^{444}). The tryptic digest of this particular region of the α1(I) chain was partitioned into three populations with different hydroxylation. The sequencing outcome in Fig 2A reflects the expected hydroxylation pattern, with the p^{444} being hydroxylated and the mass of the peptide the expected value of 1680.76. The Hyp^{444} is supported by the

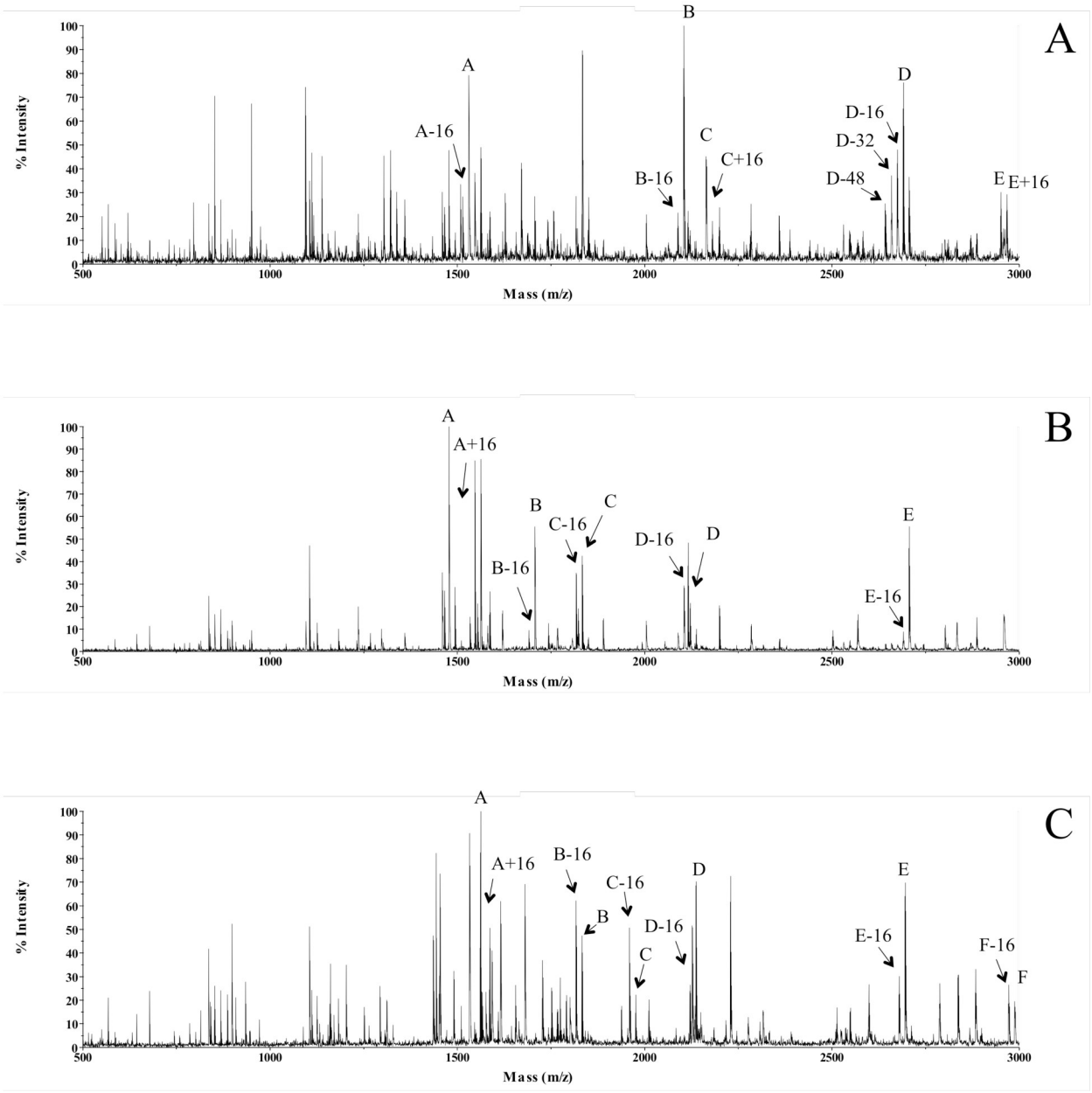

Fig 1. MALDI-TOF spectra of total trypsin digest. (A) Human collagen type III, (B) Human collagen type I, (C) Rat collagen type I. Peaks A–E in panel (A) and panel (B) and A-F in panel (C) are tryptic peptides of the corresponding collagens identified based on the agreement of their molecular weight (+1 ion) with that of the 'theoretical value' (assuming all Y-Pro as Hyp). The mass variants of 16 of each peak are labeled based on their mass differences from that of the 'theoretical value'.

https://doi.org/10.1371/journal.pone.0250544.g001

identification of a nearly complete series of y- and b-ions and most directly by the y_{10} ion with a very strong signal. The sequencing outcome of the second population indicate the peptide carries a mass variant of +16 compared to the theoretical value (Fig 2B), and the extra

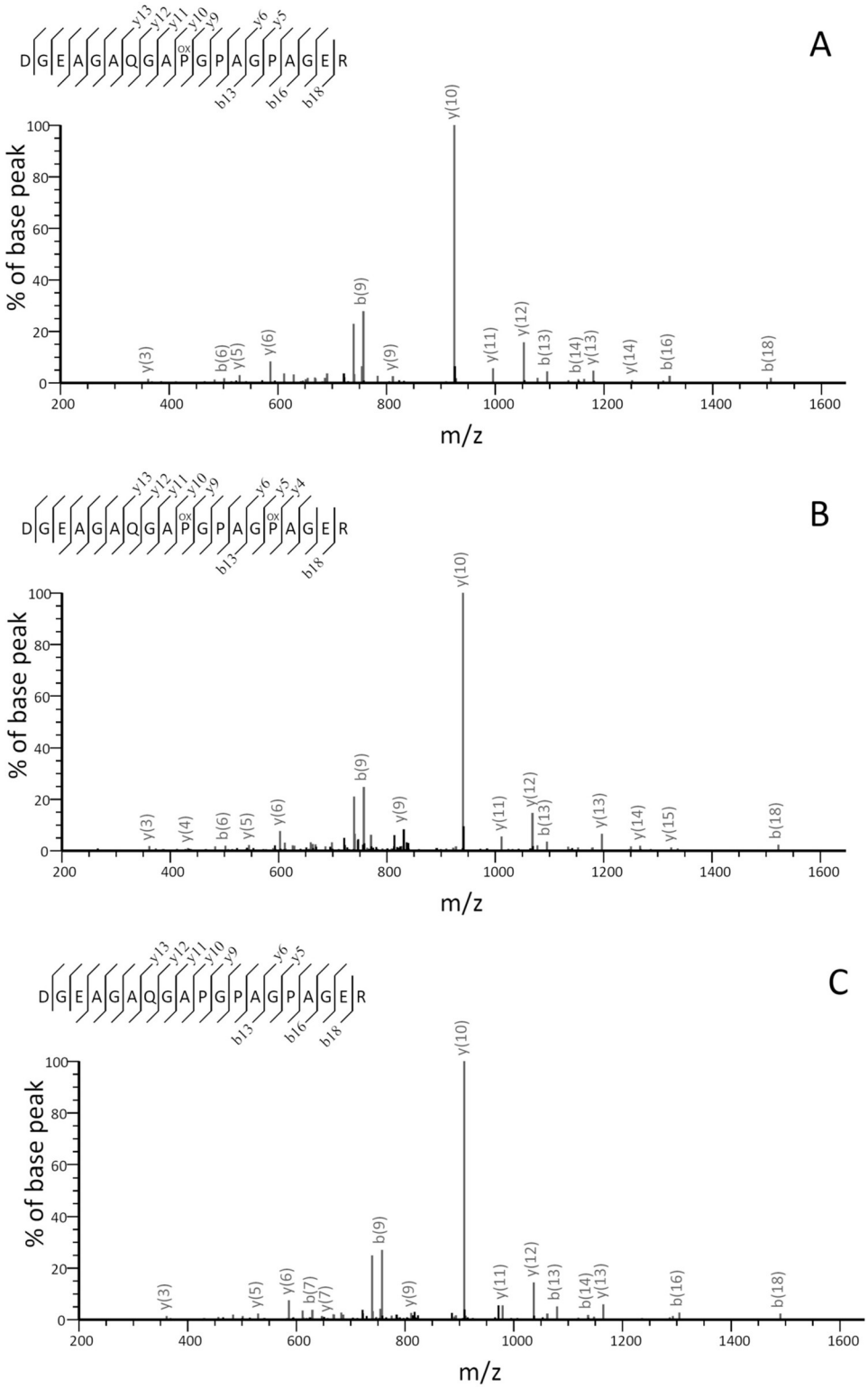

Fig 2. The MS/MS spectra of tryptic peptide ^{435}dgeagaqgapgpagpager of rat tail tendon α1(I) chain (lower case stands for the sequence from the genes). Sequencing outcome from ion 840.8832^{2+} (1680.7572^{+}) (upper panel), ion 848.8801^{2+} (1696.7521^{+}) (middle panel) and ion 832.8864^{2+} (1664.7623^{+}) (lower panel). The hydroxylation sites are shown as $\overset{ox}{P}$. For clarity, only selected ions, those most relevant to the identification of residues are labeled.

https://doi.org/10.1371/journal.pone.0250544.g002

hydroxylation is unambiguously located on the p^{449} at the X-position, directly supported by the +16 values of y_5 and y_6 ions comparing to that in Fig 2A (*i.e.*, the theoretical value). In the third case (Fig 2C) the mass has a -16 variant, and again the strong y_{10} ion, as well as the -16 mass of the y_{12} and y_{13} ions demonstrated that the p^{444} at the Y-position was not hydroxylated. The three different scenarios regarding Pro hydroxylation in this region: the one with expected Hyp^{444} in the Y-position, the one with $O_x{}^{449}$, and the one with $P_y{}^{444}$, reflects the natural variations of Pro-hydroxylation accumulated during development since this particular collagen sample is isolated from a single rat tail tendon.

The over hydroxylation and under-hydroxylation of type I collagen from rat tail tendon

The MS and MS/MS sequencing were carried out for the α1(I) and the α2(I) chains of type I collagen of rat tail tendon of a commercial sample (crt) and of a sample that is purified from a single rat tail tendon (srtt), the α1(I) and the α2(I) chains of a commercial sample of human type I collagen from placenta, and the α1(III) chain of human type III collagen from a commercial sample. The results of the peptides with unexpected hydroxylations for all five α chains of collagens are summarized in Tables 1–3 (details below); complete sequencing results of all collagen samples are given in the tables of the S1 File. All the samples are sequenced at least twice in order to evaluate the reproducibility of the results. The reproducible findings are shown in boldface in the tables. In addition to the fragmentation data, the unusual hydroxylation of the peptides in Tables 1–3 is unequivocally identified by the variations of their masses (Δm).

The results in Table 1 revealed a range of variations of the hydroxylation in both the α1 and the α2 chains of a commercial sample of the rat tail tendon type I collagen. The +16 or +32 mass confirmed the unexpected hydroxylations in 6 segments of the α1 chain at residues 145–174, 193–219, 238–252, 375–396, 658–684, and 705–725; three over-hydroxylated segments were found in the α2 chain: at residues 76–87, 145–174 and 705–725. The hydroxylation sites of these peptides except that of residues 76–87 (peptide mass 1238.6052) are assigned to a specific X-Hyp based on the fragmentation ions; the +16 mass between residues 76–87 is assigned to the C-terminal Lys since it is the only residue in that peptide that can be hydroxylated. The precise location of the extra hydroxylation of peptide mass 2605.2529 (residues 145–174) of the α2(I) chain was difficult to resolve since the terminal Pro-Lys residues were not fragmented; both are the candidates for the hydroxylation (with a +16 mass). We have tentatively assigned the hydroxylation site to be the Pro^{173} in the X-position, and the $\boldsymbol{O_x{}^{173}}$ was shown in italic in Table 1 to highlight this uncertainty. Despite lacking a clear resolution of the location of the extra hydroxylation, the same peptide was observed more than once during multiple sequencing efforts as shown in the bold face in Table 1, which is presumably related to its measurable presence in this rat tail sample. The Lys residue in the equivalent position of the α1(I) chain in the peptide of residues 145–174 was resolved unambiguously, and for multiple times by the observation of the fragment having a terminal unhydroxylated Lys. This peptide, however, was found to have an over hydroxylation site on X-Pro^{155} ($\mathbf{O_x{}^{155}}$). Similarly, the $\boldsymbol{O_x{}^{683}}$ of α1(I) chain (Table 1) and $\boldsymbol{O_x{}^{719}}$ were assigned with uncertainty because the non-fragmented terminal Pro-Lys and residues 718–725, respectively, prevented an unambiguous assignment of the unexpected hydroxylation site.

In most cases a mixed population of variable hydroxylation was observed for a particular peptide. For example, the peptide with ion mass 1306.6386 (residues 238–252) coexists with the two over-hydroxylated species: a +16 species with mass 1322.6335 and a +32 species with mass 1338.6284 carrying, respectively, one and two extra hydroxylation sites. Due to the unpredictable and complex nature of the ionization process of MS and MS/MS, it is difficult to

Table 1. Peptides of rat type I collagen with mass variants of 16 (the commercial sample).

Ion mass[a]	Δm[b]	Score[c]	Unusual Hydroxylation	Sequence of Peptide[d]
The α1(I) chain of rat collagen (commercial sample)				
1306.6386		92		^{238}GPSGPQGPSGAOGPK
1322.6335	+16	65	$\mathbf{O_x}^{245}$	^{238}GPSGPQG$\mathbf{O_x}$SGAOGPK
1338.6284	+32	58	$\mathbf{O_x}^{239}$, and $\mathbf{O_x}^{245}$	^{238}G$\mathbf{O_x}$SGPQG$\mathbf{O_x}$SGAOGPK
1328.6481	-16	74	$\mathbf{P_y}^{804}$	^{793}GFOGLOGPSGE$\mathbf{P_y}$GK
1344.6430		41		^{793}GFOGLOGPSGEOGK
1435.6812	-16	43	P_y^{297}	^{295}GEP$_y$GPSGLOGPOGER
1840.9188		50		^{705}VGPOGPSGNAGPOGPOGPVGK
1856.9137	+16	61	$\mathbf{O_x}^{707}$	^{705}VG$\mathbf{O_x}$OGPSGNAGPOGPOGPVGK
1872.9086	+32	49	$\mathbf{O_x}^{707}$, and $\boldsymbol{O_x}^{719}$	^{705}VG$\mathbf{O_x}$OGPSGNAGPO***GO$_x$OGPVGK****
1959.9519	-16	87	$\mathbf{P_y}^{273}$	^{271}GE$\mathbf{P_y}$GPAGVQGPOGPAGEEGKR
1975.9468		94		^{271}GEOGPAGVQGPOGPAGEEGKR
2014.9689		41		^{375}TGPOGPAGQDGROGPAGPOGAR
2030.9639	+16	50	O_x^{377}	^{375}TGO$_x$OGPAGQDGROGPAGPOGAR
2120.9955	-16	98	$\mathbf{P_y}^{771}$	^{757}GSOGADGPAGSOGT$\mathbf{P_y}$GPQGIAGQR
2136.9905		84		^{757}GSOGADGPAGSOGTOGPQGIAGQR
2169.0571	-16	57	P_y^{876}	^{859}GETGPAGPOGAOGAOGAP$_y$GPVGPAGK
2185.0520		45		^{859}GETGPAGPOGAOGAOGAOGPVGPAGK
2307.1364		58		^{658}GDAGPOGPAGPAGPOGPIGNVGAOGPK
2323.1313	+16	79	$\boldsymbol{O_x}^{683}$	^{658}GDAGPOGPAGPAGPOGPIGNVGAOG$\boldsymbol{O_x}$*K*
2316.0487		53		^{193}GEOGPOGPAGAAGPAGNOGADGQOGAK
2332.0436	+16	79	$\mathbf{O_x}^{206}$	^{193}GEOGPOGPAGAAG$\mathbf{O_x}$AGNOGADGQOGAK
2548.2063		101		^{145}GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	+16	54	$\mathbf{O_x}^{155}$	^{145}GNDGAVGAAG$\mathbf{O_x}$OGPTGPTGPOGFOGAAGAK
2679.2394	-16	42	$\mathbf{P_y}^{948}$	^{934}GFSGLQGPOGSOGS$\mathbf{P_y}$GEQGPSGASGPAGPR
2695.2343		127		^{934}GFSGLQGPOGSOGSOGEQGPSGASGPAGPR
The α2 (I) chain of rat collagen (commercial sample)				
1238.6052	+16	58	$\mathbf{Hyl}^{87}$	^{76}GFOGTOGLOGF$\mathbf{K_{OX}}$
1560.8129	-16	102	$\mathbf{P_y}^{891}$	^{889}GE$\mathbf{P_y}$GPAGSVGPVGAVGPR
1576.8078		59		^{889}GEOGPAGSVGPVGAVGPR
1576.7714	-32	79	$\mathbf{P_y}^{294}$, $\mathbf{P_y}^{297}$	^{292}GS$\mathbf{P_y}$GE$\mathbf{P_y}$GSAGPAGPOGLR
1592.7663	-16	58	$\mathbf{P_y}^{297}$	^{292}GSOGE$\mathbf{P_y}$GSAGPAGPOGLR
1608.7612		76		^{292}GSOGEOGSAGPAGPOGLR
1781.7937	-16	69	$\mathbf{P_y}^{273}$	^{271}GE$\mathbf{P_y}$GSAGAQGPOGPSGEEGK
1937.8948	-16	115	$\mathbf{P_y}^{273}$	^{271}GE$\mathbf{P_y}$GSAGAQGPOGPSGEEGKR
2368.0648	-16	41	$\mathbf{P_y}^{273}$	^{265}GETGNKGE$\mathbf{P_y}$GSAGAQGPOGPSGEEGK
2384.0597		60		^{265}GETGNKGEOGSAGAQGPOGPSGEEGK
1833.8977	+16	47	$\mathbf{O_x}^{707}$	^{705}TG$\mathbf{O_x}$OGPSGITGPOGPOGAAGK
2605.2529	+16	80	$\boldsymbol{O_x}^{173}$	^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAOG$\boldsymbol{O_x}$*K*
2971.5120	-16	76	P_y^{762}	^{757}GPSGEP$_y$GTTGPOGTAGPQGLLGAOGILGLOGSR
2987.5069		48		^{757}GPSGEOGTTGPOGTAGPQGLLGAOGILGLOGSR

[a] The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b] The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c] The score of Mascot search engine (see Material and Methods).

[d] Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y, respectively; those observed in more than one sequencing effort and/or in both the commercial and the srtt samples (see text) are in bold; those with uncertainties are shown in italic (see text).

* The +16 mass of the peptide of residues 718–725 (underlined) was tentatively assigned to O_x^{719} due to the lack of peptide fragmentation; the +16 mass can potentially be caused by a O_x^{722} or, with a lesser possibility (see the text), a Hyl^{725}.

https://doi.org/10.1371/journal.pone.0250544.t001

Table 2. Peptides of type I collagen from a *single* rat tail tendon with mass variants of 16.

A				
Ion mass[a]	Δ*m*[b]	Score[c]	Unusual Hydroxylation	Peptide[d]
The α1 chain of type I collagen isolated from a single rat tail tendon (srtt)				
1242.5961	-16	63	P_y^{369}	^{361}GLTGSOGSP$_y$GPDGK
1322.6335	+16	76	$\mathbf{O_x^{239}}$	^{238}G**O$_x$**SGPQGPSGAOGPK
1338.6284	+32	76	$\mathbf{O_x^{239}}$, O_x^{251}	^{238}G**O$_x$**SGPQGPSGAOGO$_x$K
1354.6233	+48	49	$\mathbf{O_x^{239}}$, $\mathbf{O_x^{245}}$, O_x^{251}	^{238}G**O$_x$**SGPQG**O$_x$**SGAOGO_x*K*
1452.7264		65		$^{-7}$SAGVSVPGPMGPSGPR
1468.7213	+16	51	M_{ox}^3	$^{-7}$SAGVSVPGPM$_{ox}$GPSGPR
1484.7162	+32	40	P^{-1*}, $\mathbf{M_{ox}^3}$	$^{-7}$SAGVSVP*GPM$_{ox}$GPSGPR
1574.8173	-16	86	P_y^{600}	^{586}GLTGPIGPOGPAGAP$_y$GDK
1590.8122		42		^{586}GLTGPIGPOGPAGAOGDK
1561.7969	+16	40	O_x^{986}	^{975}DGLNGLOGPIGO$_x$OGPR
1872.9086	+32	49	$\mathbf{O_x^{707}}$, $\mathbf{O_x^{719}}$	^{705}VG**O$_x$**OGPSGNAGPOG**O$_x$**OGPVGK
1888.9035	+48	48	$\mathbf{O_x^{707}}$, O_x^{716}, $\mathbf{O_x^{719}}$	^{705}VG**O$_x$**OGPSGNAGO$_x$OG**O$_x$**OGPVGK
1664.7623	-16	88	P_y^{444}	^{435}DGEAGAQGAP$_y$GPAGPAGER
1680.7572		96		^{435}DGEAGAQGAOGPAGPAGER
1696.7521	+16	92	O_x^{449}	^{435}DGEAGAQGAOGPAGO$_x$AGER
1800.8697	-32	105	P_y^{825}, P_y^{831}	^{817}GPOGPMGPP$_y$GLAGPP$_y$GESGR
1816.8647	-16	125	P_y^{825}, P_y^{831}, M_{ox}^{822}	^{817}GPOGPM$_{ox}$GPP$_y$GLAGPP$_y$GESGR
1803.8508	-16	77	$\mathbf{P_y^{273}}$	^{271}GE**P$_y$**GPAGVQGPOGPAGEEGK
1819.8457		82		^{271}GEOGPAGVQGPOGPAGEEGK
2105.0006	-32	78	P_y^{759}, $\mathbf{P_y^{771}}$	^{757}GSP$_y$GADGPAGSOGT**P$_y$**GPQGIAGQR
2339.1262	+32	40	O_x^{668}, $\boldsymbol{O_x^{683}}$	^{658}GDAGPOGPAGO$_x$AGPOGPIGNVGAO***GO$_x$K***
2548.2063		58		^{145}GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	+16	96	$\mathbf{O_x^{155}}$	^{145}GNDGAVGAAG**O$_x$**OGPTGPTGPOGFOGAAGAK
2663.2444	-32	96	P_y^{945}, $\mathbf{P_y^{948}}$	^{934}GFSGLQGPOGSP$_y$GS**P$_y$**GEQGPSGASGPAGPR
2679.2394	-16	105	$\mathbf{P_y^{948}}$	^{934}GFSGLQGPOGSOGS**P$_y$**GEQGPSGASGPAGPR
2695.2343		62		^{934}GFSGLQGPOGSOGSOGEQGPSGASGPAGPR
B				
The α2 chain of type I collagen isolated from a single rat tail tendon (srtt)				
868.4635		51		^{907}GPSGPQGIR
884.4585	+16	46	O_x^{908}	^{907}GO$_x$SGPQGIR
937.5102		52		^{964}GPAGPSGPIGK
953.5051	+16	51	O_x^{968}	^{964}GPAGO$_x$SGPIGK
1068.5684	-16	71	P_y^{258}	^{253}GLVGEP$_y$GPAGSK
1084.5633		70		^{253}GLVGEOGPAGSK
1175.5691	+16	79	O_x^{407}	^{397}GEAGNIGFO*GOK*
1187.6280	-16	54	P_y^{981}	^{978}SGHP$_y$GPVGPAGVR
1238.6052	+16	64	$\mathbf{Hyl^{87}}$	^{76}GFOGTOGLOGF**K$_{ox}$**
1437.7485	-16	79	P_y^{486}	^{484}GLP$_y$GEFGLOGPAGPR
1453.7434		75		^{484}GLOGEFGLOGPAGPR
1469.7383	+16	110	O_x^{494}	^{484}GLOGEFGLOGO$_x$AGPR
1490.7122		59		^{741}TGEIGASGPOGFAGEK
1506.7071	+16	67	O_x^{749}	^{741}TGEIGASGO$_x$OGFAGEK

(*Continued*)

Table 2. (Continued)

1560.8129	-16	102	**P_y^{891}**	^{889}GE**P_y**GPAGSVGPVGAVGPR
1576.8078		110		^{889}GEOGPAGSVGPVGAVGPR
2025.0373	-16	82	**P_y^{891}**	^{885}HGNRGE**P_y**GPAGSVGPVGAVGPR
2041.0322		76		^{885}HGNRGEOGPAGSVGPVGAVGPR
1576.7714	-32	76	**P_y^{294} , P_y^{297}**	^{292}GS**P_y**GE**P_y**GSAGPAGPOGLR
1592.7663	-16	85	**P_y^{294}**	^{292}GS**P_y**GEOGSAGPAGPOGLR
1608.7612		88		^{292}GSOGEOGSAGPAGPOGLR
1624.7562	+16	94	O_x^{302}	^{292}GSOGEOGSAGO$_x$AGPOGLR
1748.8674	-16	69	**P_y^{297}**	^{291}RGSOGE**P_y**GSAGPAGPOGLR
1764.8624		50		^{291}RGSOGEOGSAGPAGPOGLR
1599.8238	-16	87	P_y^{183}	^{175}GELGPVGNP$_y$GPAGPAGPR
1615.8187		105		^{175}GELGPVGNOGPAGPAGPR
1781.7937	-16	85	**P_y^{273}**	^{271}GE**P_y**GSAGAQGPOGPSGEEGK
1797.7886		94		^{271}GEOGSAGAQGPOGPSGEEGK
1817.9028		52		^{705}TGPOGPSGITGPOGPOGAAGK
1833.8977	+16	69	**O_x^{707}**	^{705}TG**O$_x$**OGPSGITGPO*G*POGAAGK
1849.8926	+32	89	**O_x^{707}, O_x^{716}**	^{705}TG**O$_x$**OGPSGITG**O$_x$**OGPOGAAGK
1865.8876	+48	97	**O_x^{707}, O_x^{716}, O_x^{719}**	^{705}TG**O$_x$**OGPSGITG**O$_x$**OG**O$_x$**OGAAGK
2589.2580		112		^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAOGPK
2605.2529	+16	103	**O_x^{173}**	^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAO*G**O$_x$**K*
2621.2478	+32	51	O_x^{164}, **O_x^{173}**	^{145}GSDGSVGPVGPAGPIGSAGO$_x$OGFOGAO*G**O$_x$**K*
3045.4257	-16	78	P_y^{552}	^{520}GPSGAOGPDGNKGEAGAVGAOGSAGASGPGGLP$_y$GER
3077.4155	+16	76	O_x^{548}	^{520}GPSGAOGPDGNKGEAGAVGAOGSAGASGO$_x$GGLOGER

[a] The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b] The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c] The score of Mascot search engine (see Material and Methods).

[d] Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y, respectively; those observed in more than one sequencing effort and/or in both the commercial and the srtt samples (see text) are in bold; those with uncertainties are shown in italic (see text).

https://doi.org/10.1371/journal.pone.0250544.t002

quantify the relative percentages of the various hydroxylated species by MS alone. In a few cases, such as the peptide with mass 2605.2529 (residues 145–174) of the α2 chain only a single population with an extra hydroxylation was sequenced. Despite two sequencing efforts, a species with the theoretical mass was not observed. This lack of detection, however, does not rule out the existence of this population in the sample *per se*. This species may fail to be sequenced to an acceptable quality either due to poor ionization and/or fragmentation, or may have failed to be identified due to unexpected post-translational modifications. One limitation of mass-spec data interpretation is the inability to draw conclusions about peptides that are not selected for fragmentation.

Concurrently, incomplete hydroxylation was found in six and four regions, respectively, of the α1(I) and the α2(I) chains, located in the regions of residues 271–291, 295–309, 757–780, 793–806, 859–884 and 934–963 of the α1(I) chain, and of residues 271–290, 292–309 (having 2 different P_y residues), 757–789 and 889–906 of the α2(I). Remarkably, among the 11 detected P_y residues in both α chains, seven of them were found in the triplet GEP$_y$.

One noticeable observation is the overlapping of the regions having unusual hydroxylation between the α1(I) and α2(I) chains. The region between residues 705 and 725 in both

Table 3. Peptides with mass variants of 16 in human collagen.

Ion mass[a] (m/z^{+1})	Δm[b]	Score[c]	Unusual Hydroxylation	Peptide[d]
A. Human collagen type I				
α1 chain				
1561.7969	+16	65	O_x^{986}	^{975}DGLNGLOGPIGO$_x$OGPR
1832.8596		78	M_{ox}^{822}, P_y^{825}	^{817}GPOGPM$_{ox}$GPP$_y$GLAGPOGESGR
1848.8545	+16	41	M_{ox}^{822}	^{817}GPOGPM$_{ox}$GPOGLAGPOGESGR
2105.0006	-16	66	**P_y^{771}**	^{757}GSOGADGPAGAOGT**P$_y$**GPQGIAGQR
2120.9955		58		^{757}GSOGADGPAGAOGTOGPQGIAGQR
2497.2066	-16	81	**P_y^{876}**	^{856}GDRGETGPAGPOGAOGAOGA**P$_y$**GPVGPAGK
2513.2015		109		^{856}GDRGETGPAGPOGAOGAOGAOGPVGPAGK
2703.2394	-16	63	**P_y^{645}**	^{619}GAOGDRGEOGPOGPAGFAGPOGADGQ**P$_y$**GAK
2149.9785		73		^{625}GEOGPOGPAGFAGPOGADGQOGAK
α2 chain				
1168.4978	-16	61	P_y^{840}	^{837}DGNP$_y$GNDGPOGR
B. Human collagen type III				
α1 chain				
949.5102	-16	45	**P_y^{981}**	^{973}GPVGPSGP**P$_y$**GK
965.5051		44		^{973}GPVGPSGPOGK
1138.5674		65		^{790}GLAGPOGMOGPR
1154.5623	+16	41	M_{ox}^{797}	^{790}GLAGPOGM$_{ox}$OGPR
1203.5827		57		^{406}GQOGVMGFOGPK
1219.5776	+16	42	*O_x^{416}*	^{406}GQOGVMGFOGO$_x$*K*
1514.7346	-16	50	P_y^{990}	^{984}DGTSGHP$_y$GPIGPOGPR
1530.7295		57		^{984}DGTSGHOGPIGPOGPR
1670.7989		42	P_y^{240}, M_{ox}^{243}	^{229}GEMGPAGIOGAP$_y$GLM$_{ox}$GAR
2088.0832	-16	94	**P_y^{777}**	^{766}GSOGAQGPOGA**P$_y$**GPLGIAGITGAR
2104.0782		137		^{766}GSOGAQGPOGAOGPLGIAGITGAR
2283.1000		79		^{667}GEGGPOGVAGPOGGSGPAGPOGPQGVK
2299.0949	+16	106	***O_x^{686}***	^{667}GEGGPOGVAGPOGGSGPA***GO$_x$OGPQGVK***[*]
2950.4653		82		^{595}GPTGPIGPOGPAGQOGDKGEGGAOGLOGIAGPR
2966.4603	+16	58	***O_x^{605}***	^{595}GPTGPIGPOG***O$_x$***AGQOGD*K*GEGGAOGLOGIAGPR

[a] The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b] The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c] The score of Mascot search engine (see Material and Methods).

[d] Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y; those observed in more than one sequencing effort are in bold; those with uncertainties are shown in italic (see text).

[*] The +16 mass of the peptide of residues 685–693 (underlined) was tentatively assigned to O_x^{686} due to the lack of peptide fragmentation.

https://doi.org/10.1371/journal.pone.0250544.t003

α chains contain O_x residues: O_x^{707} and O_x^{719} in α1(I) chain and O_x^{707} in α2(I) chain. Similarly, some of the P_y residues appear located in similar regions of both α chains as well: P_y^{273} between residues 271–291, P_y^{294} and P_y^{297} between residues 291–309, P_y^{762} and P_y^{771} between residues 757–789. Combining the observations of different peptides from both α chains, the region of residues 271–309 in both α1(I) and α2(I) chains stands out as a particularly poorly hydroxylated region, missing two to three expected Hyp residues in the Y-positions of each α chain.

The finding of such a wide range of variations in hydroxylation of the type I chain was rather unexpected. The purity and the purification procedures of this commercial sample were called into question. In order to get a better understanding of the origin of the heterogeneity we purified the type I collagen from a single rat tail tendon (srtt). Interestingly, the sequencing result of this srtt sample turns out to be remarkably similar (Table 2). The 2 observed O_x residues in the α2(I) of the commercial sample and all but two (O_x^{206} and O_x^{377}) in the α1(I) chain (Table 1) were reproduced in the srtt sample. Similarly, more than half of the P_y residues found in the commercial sample were also observed in the srtt sample. This srtt sample appeared to be particularly over-hydroxylated having 11 O_x residues in each α chain. The content of P_y is also higher: 10 and 9 P_y residues, respectively, were found in the α1(I) and α2(I) chains. The noticeably more heterogeneous hydroxylation pattern of this srtt sample, especially that for the α2(I) chain, may relate to the better overall sequencing outcome of the sample reflected, in part, by the better sequence coverage of this α2 chain (Fig 3 and S1-S6 Tables in S1 File). The identified O_x residues of the α1(I) chain of the srtt sample include the well-known 3-Hyp O_x^{986}; this section of the α1(I) chain of the commercial sample was not sequenced. Another interesting peptide with an additional Hyp of interest is the seven-residue peptide from the N-telopeptide region preceding the triple helical domain (ion mass 1452.7264). The Met residue in this fragment appears to be oxidized based on the detection of M_{ox} with neutral loss. In addition to the M_{ox}, the Pro in the N-telopeptide appears to be hydroxylated (the P^*). Since this Pro precedes a Gly residue, which characterizes the canonical hydroxylation site of the prolyl-hydroxylase (C-P4H), its hydroxylation, although never reported before, probably does not come as too much of a surprise.

The regions identified in the commercial sample where the α1(I) and α2(I) chains are over-hydroxylated, residues 238–252 and residues 705–725 (Table 1), also have multiple O_x in this srtt sample. The region of residues 705–725 revealed a particularly varied hydroxylation pattern having 1 to 3 O_x residues in both chains. By comparing to the commercial sample, another highly variable region stands out: residues 238–252 of α1(I) chain having up to 2 and 3 O_x residues in the commercial sample and the srtt sample, respectively. The over-hydroxylation, however, is not seen for the equivalent region of the α2(I) chain because of the non-homologous sequences: none of the equivalent X-residues of the α2(I) chain where an O_x is observed in the α1(I) chain is Pro. The poorly hydroxylated region of residues 271–309 is also under-hydroxylated in this srtt sample, lacking 1 and 3 expected Y-Hyps, respectively, in the α1(I) and the α2 (I) chains. Other P_y residues that are frequently observed in both samples are P_y^{771} and P_y^{948} of the α1(I) chain, and P_y^{891} of the α2(I) chain.

In fact, the sequenced peptides of the commercial sample in Table 1 appear almost as a subset of that included in Table 2A and 2B of the srtt sample. Thus, on this account, the commercial samples are quite representative of the averaged features of PTMs of collagens from the type I collagen of the rat tail tendon. All together combining the results of the two samples of type I collagen we have identified a total of 13 O_x in the α1 (I) chain and 12 in the α2(I) chains, also 13 and 10 P_y residues, respectively, in the α1(I) and α2(I) chains. The abnormal hydroxylation sites detected from *both* samples are mapped out on the sequences of the α-chains arranged in D-periodicity in Fig 3. The sites of O_x appear to be scattered rather uniformly throughout the α-chains except the two identified regions of highly variable hydroxylation patterns (HVRs): the N-terminal highly variable region (N-HVR) of residues 238–252 of the α1(I) chain, and the C-terminal highly variable region (C-HVR) of residues 705–725 of both α chains. The O_x^{155} and O_x^{683} of the α1(I) chain are the only unexpected hydroxylations outside the HVRs that have been observed multiple times in both samples. The P_y residues also seem to cluster: in addition to residues 271–302 mentioned above, regions of residues 941–950 and 821–840 and 757–780 of the α1(I) chain all have multiple P_y residues (Fig 3 and Tables 1 and 2A and 2B).

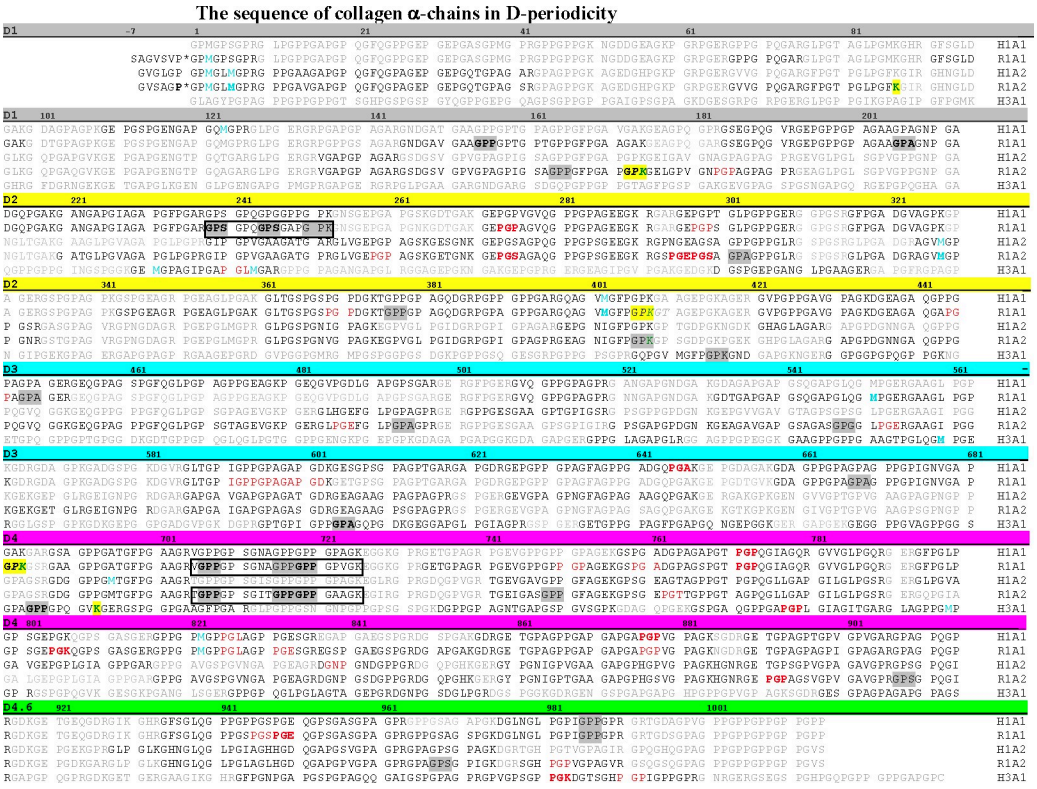

Fig 3. The mapping of individual unexpected hydroxylation sites on the α chains of collagen. Sequences of the α_1(I) and α_2(I) chains of human (H1A1 and H1A2, respectively), the α_1(I) and the α_2(I) chains of rat (R1A1 and R1A2, respectively), and the α1(III) chain of human (H3A1) were arranged by the *D*-periodicity according to Di Lullo *et al* [54]: the 4 *D*-periods are highlighted by a colored bar of grey, yellow, cyan, and magenta, respectively; the 0.6 *D* is marked by the colored bar of green. The Gly-X-Y triplets including an O_x are in grey highlight. The P_y residues in the tripeptide unit of Y-Gly-X are shown in red in order to reflect the potential connection with the enzyme selectivity of C-P4H (see text). The entire segment of the three highly variable regions (N- or C-HVR, see text) with multiple O_x residues are boxed. The hydroxylated proline in the telopeptide is P*. Hydroxylysines and the Gly-Pro-Lys tripeptide where the hydroxylation could not be precisely located between the Pro-Lys residues at the C-terminus of a peptide (see text) are highlighted in yellow with Lys in green font. The oxidized methionines are in blue colored font. In all cases the PTMs observed from more than one detection/sample preparation were shown in bold font. Not sequenced regions are in faint grey. The amino acid sequence of the five collagen α chains were adapted from the UniProt database.

https://doi.org/10.1371/journal.pone.0250544.g003

Human collagen type I and type III

The detected unusual hydroxylation sites of human collagen type I and type III are summarized in Table 3. Only the primary 3-Hyp at position 986 (O_x^{986}) of α1(I) was found as the X-Hyp in human type I collagen (Table 3A). Despite multiple sequencing efforts the peptide containing Pro^{707} of α2(I), one of the class-2 X-Hyp reported by Eyre and colleagues, was not sequenced. The Pro^{707} of α1(I) was sequenced but was found not hydroxylated in spite of the nearly identical amino acid sequences in this region between the two α-chains (Fig 3). The residue Met^{822} appears to be oxidized (Table 3A). The oxidation of Met is not a regular post-translational modification but an oxidation event usually found in cells under stress; it can also occur with sample handling [52, 55, 56]. A few cases of incomplete hydroxylations were observed for both the α1 (I) chain and the α2(I) chain. In general, the sequencing of the human type I collagen sample appeared to be rather clean, with only a low degree of unexpected modifications.

Additional hydroxylation was seen in three regions of the α1(III) chain of the type III collagen (Table 3B): peptides of residues 406–417 (mass 1219.5776), 595–627 (peptide mass

2966.6403), and 667–693 (mass 2299.0949). The peptide with mass 2966.4603 contains an internal Lys^{612}, the +16 mass was tentatively assigned to the hydroxylation of O_x^{605} due to the lack of complete peptide fragmentation between O_x^{605} and Lys^{612}. Similarly, the +16 mass was tentatively assigned to O_x^{686} for the lack of fragmentation between residues 685–693. The peptide 406–417 (mass 1219.5776) carries the Met^{411}, which can potentially be oxidized with a mass increase of 16; the fragmentation data has ruled out this possibility. Overall four P_y residues were detected in α1(III) samples, and the P_y^{981} was seen with a very strong signal in every sequencing outcome.

In summary, the MS/MS sequencing results are mapped out on the sequences of the five α-chains arranged in *D*-periodicity in Fig 3. Using the rather stringent sequencing criteria outlined in Materials and Methods the sequence coverage is about 35% for type III collagen and around 56% for the α_1(I) and 46% for the α_2(I) chains of human type I collagen, and about 62% and 67% for α_1(I) and α_2(I) chains, respectively, of rat tail tendon type I collagen combining both samples. Some observations of the O_x are consistent between the two different rat tail tendon samples, such as those in the HVRs; the others appeared sporadic. Most of the over-hydroxylations seen in the rat tail tendon type I collagen are not present in human placental type I collagen. The consistent observations between the human type I collagen and that of the rat tail tendon include the Hyp^{986} of the α1 chain and P_y resides at positions 771 and 876 of the α1(I) chain.

Discussion

By carrying out this study of the selected collagen samples we are hoping to gain a better understanding of the variations in the hydroxylation of fibrillar collagen in MS studies. Because of the high sensitivity of the MS and MS/MS approach, observing unusual hydroxylation of collagen proves to be a common event. Using a standard protein mass-spec sequencing technique we have detected unusual hydroxylation at several levels in rat type I collagen, human type I collagen and human type III collagen. The variations of the hydroxylation were supported by the spectrometry data for both the fragmented ions and the overall mass of the tryptic peptides. The over-hydroxylation was largely attributed to the hydroxylation of Pro in the X-positions, which is especially prevalent in both the α1(I) and the α2(I) chains of the type I collagen of rat tail tendon. The heterogeneity in the hydroxylation of human collagen type I and type III is much lower, reflecting the variations in enzyme selectivity of the hydroxylase among different species and/or tissues. As expected, most of the hydroxylated proline residues in the X-position are detected as a mixture; some may be present at a relatively low level, while others, as those in the highly variable regions (HRVs), are more prevalent and representative.

The repeated sequencing outcomes of the same collagen sample often carry high levels of variations as shown in Tables 1–3; detections of about half of the O_x and P_y residues are seen in multiple sequencing efforts (in boldface), while that of the others are less reproducible. In fact, the variations in the sequencing results of the two very different rat tail tendon samples are not in any way more substantial than that of the repeated sequencings of the same collagen samples. Such varied outcomes reflected the complex and unpredictable nature of the ionization process of MS and sample handling [57]. Each sequencing outcome often represents no more than a single sampling of a population consisting of heterogeneous modifications. The unpredictable ionization process is one of the major concerns for quantitative estimation of the populations of the sequenced peptides using MS/MS, especially when the sample is heterogeneous and the scope of the PTMs of the protein is not fully characterized. Sequence coverage will also affect the detection of PTMs, and this may be the reason that the canonical $3Hyp^{986}$ of the α1(I) chain was detected only once among multiple sequencing attempts of samples from

human placenta and from srtt; this tryptic peptide containing position 986 was not sequenced at all in the commercial rat tail sample. Protocols using multiple proteases will result in a better sequence coverage, especially in cases like type III collagen where the tryptic peptides are often either too large or too small for reliable MS/MS results. On the other hand, despite the low sequence coverage, several unusual hydroxylations were observed rather consistently (Table 3 and Fig 3). The observations of the O_x residues, in the highly variable regions of the rat tail tendon type I collagen is quite robust and consistent even among samples prepared from different sources.

The over hydroxylation observed in the C-HVR of the α1(I) and α2(I) chains of type I collagen in rat tail tendon is in keeping with the unusually high 3Hyp content of this collagen [58]. The amino acid composition analysis estimated three to four 3Hyp in each of the α1(I) and the α2(I) chains of the rat tail tendon type I collagen, compared to only one 3Hyp, the Hyp^{986}, in the rat α1(I) chain of type I collagen from bones or skin. The O_x^{707}, O_x^{716} and O_x^{719} of the α2(I) chains were subsequently identified as 3Hyp by N-terminal sequencing [3]. The C-HVR also includes the location of the 'class-2' $3Hyp^{707}$ in the human α2(I) chain observed previously [8]. Unfortunately this segment of the human α2(I) chain was not sequenced in our study despite repeated attempts; the same region in the human α1(I), which has the same amino acid sequence as that in the α2(I) chain, was sequenced, but no X-Hyp was found. The over-hydroxylation in the N-HVR of residues 238–252 of rat tendon α1(I) has never been reported before. Multiple O_x residues in this region in both samples of the rat tendon collagen were observed with reproducible results. Interestingly, the amino acid sequences of the two highly variable regions share limited homology; they are also rather different from the sequence surrounding the $3Hyp^{986}$. The two O_x residues of the N-HVR, O_x^{239} and O_x^{245}, appeared in the peptide triad of GO_xS, while the O_x^{707}, O_x^{716} and O_x^{719} of C-HVR in both α1(I) and α2(I) chains are in the more common GO_xO moiety. The GOA and GOS are two moieties identified for X-Hyp of type V collagen [9]. It is also worth noting that, similar to the 3Hyp residues identified by Eyre and colleagues, the two HVRs of rat tail tendon type I collagen are located exactly a 2*D*-period apart (Fig 3), although the significance of it remains to be evaluated [8]. No further effort was made to confirm the 3Hyp identity of the identified O_x in this study. While most of the newly identified X-Hyp residues have been confirmed to be 3Hyp, at least in one occasion an X-Hyp was later confirmed to be a 4R-Hyp [59].

The unhydroxylated Pro residues in the Y-position appear to be more common than X-Hyp among all α chains, with the highest content seen in the rat tail tendon type I collagen. Most of the detected Y-Pro residues are present as a mixed population having varied occupancies. Combining all the five α chains, 32 P_y residues were observed in 28 peptides. It is tempting to postulate the region of residues 273–302 of rat tail tendon type I collagen, where up to 5 P_y residues were found within a short stretch of 30-residues, to have unique conformational dynamics, since a Pro in the Y-position is known to significantly destabilize the triple helix compared to a Hyp [60]. The real impact will, of course, depend on the percent of occupancy in these sites.

While incomplete hydroxylation has been known for some time, the site-specific data and the sequence motif of the missed-hydroxylations have not been reported before. The sequence information of the P_y residues may relate to the substrate selectivity of the prolyl-4-hydroxylase (C-P4H). Studies using short peptides established that the enzyme recognizes Pro-Gly-Xaa triplets during hydroxylation, where the Pro is the residue to be hydroxylated, and the selection of the Pro in a Y-position is affected by the conformation around the -Gly-Xaa residues [61–63]. The hydroxylation takes place on the nascent polypeptide chains before the formation of the triple helix. Despite the higher than normal content of the Pro residues, the unfolded α-chains of collagen are not known to assume any well-defined conformation, although isolated

segments may temporarily adapt to polyproline II (PP II) like or β-turn like ϕ and φ angles. Specifically, the type II β-turn bent between the -Gly-Xaa was considered to favor the binding of C-P4H and thus, the hydroxylation of the Pro proceeding the Gly; while the PPII conformation in -Gly-Xaa was considered inhibitory [62, 63]. Residues Ala, Leu, Ile, and Phe in the position of Xaa were found to favor a β-turn around the Gly, and a Pro favors a PPII ϕ and φ angles [63]. Our finding here appears to reflect this conformational preference of C-P4H in vivo. If we consider the P_y as a *miss* of the C-P4H in selecting a Pro in a Y-position for hydroxylation, the hydroxylation action appears to be particularly slippery in the sequence context of Pro-Gly-Pro. Fourteen out of the 32 P_y residues identified in the 5 α chains are in the P_y-Gly-Pro moiety. On the other hand, this high occurrence of P_y-Gly-Pro moiety may simply reflect the higher frequencies of genomic sequence pro-gly-pro in fibrillar collagen. The other frequent *misses* include P_y in Pro-Gly-Glu (5/32), Pro-Gly-Ser (3/32) triplets, and Pro-Gly-Leu (3/32) triplet. Among the identified P_y residues, the P_y^{771} of α1(I) in a P_y-Gly-Pro moiety is the only one that is identified in both human and the two rat tail tendon samples. The P_y^{876}, also in P_y-Gly-Pro moiety, of α1(I) was detected in human and the commercial sample of rat tail tendon, but not in the srtt sample. There also appears to be an overrepresentation of P_y residues in a GEP_y tripeptide in rat type I collagen: 8 out of the 23 identified P_y residues are in a GEP moiety. It is not clear if the Glu preceding the Pro in the Y-position affects the selectivity of C-P4H in rat tail tendon. Other more common sequence motifs for P_y are GAP_y, GPP_y and GSP_y. It is also unclear if the missed hydroxylation of these residues has any functional roles for collagen.

Conclusions

The high sensitivity of MS/MS sequencing has revealed a subpopulation of collagen that bares unexpected hydroxylated Pro in the X-positions, and unhydroxylated Pro in the Y-positions. The observations of the unexpected modifications, especially those present in a low population, are often inconsistent between samples due to the limit of the sensitivity of the technique and/or the tissue/organism dependent variations of the Pro-hydroxylation reaction. The detection of some modifications such as those in the HVRs of rat tail tendon type I collagen, however, appears to be quite robust and can be used as a biomarker for general applications using MS/MS. A more thorough understanding of the dynamics of the specific PTM of 3-Hyp in the X-position and its role in epigenetic regulation will require a knowledge base that is broad enough to reflect the statistical nature of both the variations of the PTMs in different tissues and organisms, and the reproducibility of their detections by MS/MS.

Supporting information

S1 File. The complete sequencing results.
(DOCX)

S1 Fig. DSD-PAGE of type I and type III collagen. A: lane 1 –human collagen type III, lane 2 = molecular marker (Sigma). B: lane 1 = rat tail tendon collagen type I, lane 2 = molecular marker (Sigma). Gels were stained with coomassie blue.
(PDF)

S2 Fig. Uncropped gel picture of collagen type III. From left: Lane 1 = molecular marker, Lane 2 = type I collagen, Lane 3 = type I collagen, Lane 4 = collagen type III, Lane 5 = marker, Lane 6 = collagen type I (rat), Lane 7 = collagen type I (human). Lanes 4 and 5 are used in S1 Fig.
(JPG)

S3 Fig. Uncropped gel of collagen type I. From left: Lane 1 = high range molecular marker, Lane 2 = molecular marker, Lane 3 = collagen type III, Lane 4 = BSA, Lane 5 = low range molecular marker. The rest lanes of the gel were empty. Lanes 2 and 3 are used in S1 Fig.
(JPG)

Acknowledgments

The authors are indebted to Dr. Sergey Leikin, and Dr. Elena N. Makareeva for their help preparing rat tail collagen, Dr. Milica Tesic Marks for some sequencing and data processing, Dr. Henrik Molina for Discoverer software assistance and the overall guidance on mass-spec technology. We would also like to thank Drs. Rebecca Strawn, James San Antonio and late Adel Bosky for critical reading of this and a previous version of the manuscript.

Author Contributions

Conceptualization: Yujia Xu.

Data curation: Michele Kirchner.

Formal analysis: Michele Kirchner, Haiteng Deng, Yujia Xu.

Supervision: Haiteng Deng, Yujia Xu.

Writing – original draft: Yujia Xu.

Writing – review & editing: Michele Kirchner, Yujia Xu.

References

1. van Huizen N.A., et al., Collagen analysis with mass spectrometry. Mass Spectrom Rev, 2020. 39(4): p. 309–335. https://doi.org/10.1002/mas.21600 PMID: 31498911
2. Henkel W. and 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): p. 4269–89. https://doi.org/10.1021/pr070318r PMID: 17939700
3. Taga Y., et al., Developmental Stage-dependent Regulation of Prolyl 3-Hydroxylation in Tendon Type I Collagen. J Biol Chem, 2016. 291(2): p. 837–47. https://doi.org/10.1074/jbc.M115.686105 PMID: 26567337
4. Eyre D.R., Paz M.A., and Gallop P.M., Cross-linking in collagen and elastin. Annu Rev Biochem, 1984. 53: p. 717–48. https://doi.org/10.1146/annurev.bi.53.070184.003441 PMID: 6148038
5. Eyre D.R., et al., A novel 3-hydroxyproline (3Hyp)-rich motif marks the triple-helical C terminus of tendon type I collagen. J Biol Chem, 2011. 286(10): p. 7732–6. https://doi.org/10.1074/jbc.C110.195768 PMID: 21239503
6. Eyre D.R. and Wu J.-J., Collagen Cross-Links. Top Curr Chem, 2005. 247: p. 207–229.
7. Fernandes R.J., et al., A role for prolyl 3-hydroxylase 2 in post-translational modification of fibril-forming collagens. J Biol Chem, 2011. 286(35): p. 30662–9. https://doi.org/10.1074/jbc.M111.267906 PMID: 21757687
8. Weis M.A., et al., Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. J Biol Chem, 2010. 285(4): p. 2580–90. https://doi.org/10.1074/jbc.M109.068726 PMID: 19940144
9. Yang C., et al., Comprehensive mass spectrometric mapping of the hydroxylated amino acid residues of the alpha1(V) collagen chain. J Biol Chem, 2012. 287(48): p. 40598–610. https://doi.org/10.1074/jbc.M112.406850 PMID: 23060441
10. Buckley M., et al., Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun Mass Spectrom, 2009. 23(23): p. 3843–54. https://doi.org/10.1002/rcm.4316 PMID: 19899187
11. Engel J. and Bachinger H.P., Structure, stability and folding of the collagen triple helix. Top Curr Chem, 2005. 247: p. 7–33.

12. Birk D.E. and Bruckner P., Collagen suprastructures. Top Curr Chem, 2005. 247: p. 185–205.
13. Cabral W.A., et al., Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nature genetics, 2007. 39(3): p. 359–65. https://doi.org/10.1038/ng1968 PMID: 17277775
14. Marini J.C., et al., Osteogenesis imperfecta. Nat Rev Dis Primers, 2017. 3: p. 17052. https://doi.org/10.1038/nrdp.2017.52 PMID: 28820180
15. Maity P.P., et al., Isolation and mass spectrometry based hydroxyproline mapping of type II collagen derived from Capra hircus ear cartilage. Commun Biol, 2019. 2: p. 146. https://doi.org/10.1038/s42003-019-0394-6 PMID: 31044171
16. Taga Y., et al., Stable isotope-labeled collagen: a novel and versatile tool for quantitative collagen analyses using mass spectrometry. J Proteome Res, 2014. 13(8): p. 3671–8. https://doi.org/10.1021/pr500213a PMID: 24953783
17. Fang M., et al., Collagen as a double-edged sword in tumor progression. Tumour Biol, 2014. 35(4): p. 2871–82. https://doi.org/10.1007/s13277-013-1511-7 PMID: 24338768
18. Zhen E.Y., et al., Characterization of metalloprotease cleavage products of human articular cartilage. Arthritis Rheum, 2008. 58(8): p. 2420–31. https://doi.org/10.1002/art.23654 PMID: 18668564
19. Szpak P., Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis. J Archaeol Sci, 2011. 38(12): p. 3358–3372.
20. Kielty C.M., Hopkinson I., and Grant M.E., The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. in Royce P. M., and Steinmann B. (Eds.), Connective Tissue and Its Heritable Disorders, pp. 103–147, Wiley-Liss, New York, 1993: p. 103–147.
21. Piez K.A., Primary Structure. in Ramachandran G. N. and Reddi A. H. (Eds), Biochemistry of Collagens, Plenum Press, New York and London. 1976. p. 1–44.
22. Bornstein P. and Piez K.A., The nature of the intramolecular cross-links in collagen. The separation and characterization of peptides from the cross-link region of rat skin collagen. Biochemistry, 1966. 5(11): p. 3460–73. https://doi.org/10.1021/bi00875a012 PMID: 5972327
23. Kivirikko K.I., et al., Further hydroxylation of lysyl residues in collagen by protocollagen lysyl hydroxylase in vitro. Biochemistry, 1973. 12(24): p. 4966–71. https://doi.org/10.1021/bi00748a023 PMID: 4761977
24. Ogle J.D., Arlinghaus R.B., and Lgan M.A., 3-Hydroxyproline, a new amino acid of collagen. J Biol Chem, 1962. 237: p. 3667–73. PMID: 13939597
25. Prockop D.J., et al., Intracellular Steps in the Biosynthesis of Collagen. in Ramachandran G. N. and Reddi A. H. (Eds), Biochemistry of Collagens, Plenum Press, New York and London, 1976: p. 163–274.
26. Privalov P.L., Tictopulo E.I., and Tischenko V.M., Stability and Mobility of the Collagen Structure. J Mol Biol, 1979. 127: p. 203–216. https://doi.org/10.1016/0022-2836(79)90240-7 PMID: 430563
27. Holmes D.F. and Kadler K.E., The 10+4 microfibril structure of thin cartilage fibrils. Proc Natl Acad Sci U S A, 2006. 103(46): p. 17249–54. https://doi.org/10.1073/pnas.0608417103 PMID: 17088555
28. Ruggiero F., et al., Triple Helix Assembly and Processing of Human Collagen Produced in Transgenic Tobacco Plants. FEBS Letters, 2000. 469: p. 132–136. https://doi.org/10.1016/s0014-5793(00)01259-x PMID: 10708770
29. Hulmes D.J., et al., Analysis of the primary structure of collagen for the origins of molecular packing. J Mol Biol, 1973. 79(1): p. 137–48. https://doi.org/10.1016/0022-2836(73)90275-1 PMID: 4745843
30. Berg R.A. and Prockop D.J., The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem Biophys Res Commun, 1973. 52(1): p. 115–20. https://doi.org/10.1016/0006-291x(73)90961-3 PMID: 4712181
31. Sakakibara S., et al., Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. Biochim Biophys Acta, 1973. 303(1): p. 198–202. https://doi.org/10.1016/0005-2795(73)90164-5 PMID: 4702003
32. Engel J., et al., The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)n and (L-Pro-L-Hyp-Gly)n. Biopolymers, 1977. 16(3): p. 601–622. https://doi.org/10.1002/bip.1977.360160310 PMID: 843606
33. Persikov A.V., et al., Amino Acid Propensities for the Collagen Triple-Helix. Biochemistry, 2000. 39: p. 14960–14967. https://doi.org/10.1021/bi001560d PMID: 11101312
34. Burjanadze T.V., New analysis of the phylogenetic change of collagen thermostability. Biopolymers, 2000. 53(6): p. 523–8. https://doi.org/10.1002/(SICI)1097-0282(200005)53:6<523::AID-BIP8>3.0.CO;2-7 PMID: 10775067

35. Saito M. and Marumo K., Effects of Collagen Crosslinking on Bone Material Properties in Health and Disease. Calcif Tissue Int, 2015. 97(3): p. 242–261. https://doi.org/10.1007/s00223-015-9985-5 PMID: 25791570

36. Heikkinen J., et al., Lysyl Hydroxylase 3 Is a Multifunctional Protein Possessing Collagen Glucosyltransferase Activity. J Biol Chem, 2000. 275(46): p. 36158–36163. https://doi.org/10.1074/jbc.M006203200 PMID: 10934207

37. Sricholpech M., et al., Lysyl Hydroxylase 3 Glucosylates Galactosylhydroxylysine Residues in Type I Collagen in Osteoblast Culture. J Biol Chem, 2011. 286(11): p. 8846–8856. https://doi.org/10.1074/jbc.M110.178509 PMID: 21220425

38. Valli M., et al., Deficiency of CRTAP in non-lethal recessive osteogenesis imperfecta reduces collagen deposition into matrix. Clin genet, 2012. 82(5): p. 453–9. https://doi.org/10.1111/j.1399-0004.2011.01794.x PMID: 21955071

39. Marini J.C., et al., Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development. Cell cycle, 2007. 6(14): p. 1675–81. https://doi.org/10.4161/cc.6.14.4474 PMID: 17630507

40. Lothrop A.P., Torres M.P., and Fuchs S.M., Deciphering post-translational modification codes. FEBS letters, 2013. 587(8): p. 1247–57. https://doi.org/10.1016/j.febslet.2013.01.047 PMID: 23402885

41. Chait B.T., Mass spectrometry in the postgenomic era. Annu Rev Biochem, 2011. 80: p. 239–46. https://doi.org/10.1146/annurev-biochem-110810-095744 PMID: 21675917

42. Lange V., et al., Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol, 2008. 4: p. 222. https://doi.org/10.1038/msb.2008.61 PMID: 18854821

43. Hunt D.F., et al., Protein sequencing by tandem mass spectrometry. Proc Natl Acad Sci U S A, 1986. 83(17): p. 6233–7. https://doi.org/10.1073/pnas.83.17.6233 PMID: 3462691

44. Huang Y., et al., Dissociation behavior of doubly-charged tryptic peptides: correlation of gas-phase cleavage abundance with ramachandran plots. J Am Chem Soc, 2004. 126(10): p. 3034–5. https://doi.org/10.1021/ja038041t PMID: 15012117

45. Vaisar T. and Urban J., Probing the proline effect in CID of protonated peptides. J Mass Spectrom, 1996. 31(10): p. 1185–7. https://doi.org/10.1002/(SICI)1096-9888(199610)31:10<1185::AID-JMS396>3.0.CO;2-Q PMID: 8916427

46. Makareeva E., et al., Molecular mechanism of alpha 1(I)-osteogenesis imperfecta/Ehlers-Danlos syndrome: unfolding of an N-anchor domain at the N-terminal end of the type I collagen triple helix. J Biol Chem., 2006. 281(10): p. 6463–70. https://doi.org/10.1074/jbc.M511830200 PMID: 16407265

47. Clough T., et al., Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. BMC Bioinformatics, 2012. 13 Suppl 16(Suppl 16): p. S6. https://doi.org/10.1186/1471-2105-13-S16-S6 PMID: 23176351

48. Fenyö D. and Beavis R.C., A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem, 2003. 75(4): p. 768–74. https://doi.org/10.1021/ac0258709 PMID: 12622365

49. Lee H.Y., et al., Refinements of LC-MS/MS Spectral Counting Statistics Improve Quantification of Low Abundance Proteins. Sci Rep, 2019. 9(1): p. 13653. https://doi.org/10.1038/s41598-019-49665-1 PMID: 31541118

50. Serang O. and Noble W., A review of statistical methods for protein identification using tandem mass spectrometry. Stat Interface, 2012. 5(1): p. 3–20. https://doi.org/10.4310/sii.2012.v5.n1.a2 PMID: 22833779

51. Steen H. and Mann M., The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol, 2004. 5(9): p. 699–711. https://doi.org/10.1038/nrm1468 PMID: 15340378

52. Bern M., Saladino J., and Sharp J.S., Conversion of methionine into homocysteic acid in heavily oxidized proteomics samples. Rapid Commun Mass Spectrom, 2010. 24(6): p. 768–72. https://doi.org/10.1002/rcm.4447 PMID: 20169556

53. Molony M.S., et al., Hydroxylation of Lys residues reduces their susceptibility to digestion by trypsin and lysyl endopeptidase. Anal Biochm, 1998. 258(1): p. 136–7. https://doi.org/10.1006/abio.1997.2513 PMID: 9527859

54. Di Lullo G.A., et al., Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. J Biol Chem, 2002. 277(6): p. 4223–31. https://doi.org/10.1074/jbc.M110709200 PMID: 11704682

55. Guan Z., Yates N.A., and Bakhtiar R., Detection and characterization of methionine oxidation in peptides by collision-induced dissociation and electron capture dissociation. J Am Soc Mass Spectr, 2003. 14(6): p. 605–13. https://doi.org/10.1016/S1044-0305(03)00201-0 PMID: 12781462

56. Houde D., et al., Determination of protein oxidation by mass spectrometry and method transfer to quality control. J Chromatogr A, 2006. 1123(2): p. 189–98. https://doi.org/10.1016/j.chroma.2006.04.046 PMID: 16716331

57. Cox J.T., et al., On the ionization and ion transmission efficiencies of different ESI-MS interfaces. J Am Soc Mass Spectr, 2015. 26(1): p. 55–62. https://doi.org/10.1007/s13361-014-0998-5 PMID: 25267087

58. Piez K.A., Eigner E.A., and Lewis M.S., The Chromatographic Separation and Amino Acid Composition of the Subunits of Several Collagens. Biochemistry, 1963. 2((1)): p. 58–66.

59. van Huizen N.A., et al., Identification of 4-Hydroxyproline at the Xaa Position in Collagen by Mass Spectrometry. J Proteome Res, 2019. 18(5): p. 2045–2051. https://doi.org/10.1021/acs.jproteome.8b00930 PMID: 30945869

60. Burjanadze T.V. and Bezhitadze M.O., Presence of a thermostable domain in the helical part of the type I collagen molecule and its role in the mechanism of triple helix folding. Biopolymers, 1992. 32(8): p. 951–6. https://doi.org/10.1002/bip.360320806 PMID: 1420979

61. Bhatnagar R.S. and Rapaka R.S., Synthetic Polypeptide Models of Collagen: Synthesis and Applications. in Ramachandran G. N. and Reddi A. H. (Eds), Biochemistry of Collagens, Plenum Press, New York and London, 1976: p. 479–523.

62. Prockop D.J. and Kivirikko K.I., Effect of polymer size on the inhibition of protocollagen proline hydroxylase by polyproline II. J Biol Chem, 1969. 244(18): p. 4838–42. PMID: 5824558

63. Brahmachari S.K. and Ananthanarayanan V.S., Beta-turns in nascent procollagen are sites of post-translational enzymatic hydroxylation of proline. Proc Natl Acad Sci U S A, 1979. 76(10): p. 5119–23. https://doi.org/10.1073/pnas.76.10.5119 PMID: 228279