Identification of collagen types in tissues using HPLC-MS/MS

A method for the determination and quantification of collagen types I–V in rat tissues has been developed. This method is based on collagen fragmentation by cyanogen bromide followed by trypsin digestion. After that, HPLC-MS/MS (HPLC coupled to an IT mass spectrometer) analyses of the resulting peptide mixtures (peptide maps) were performed. Specific peptides for each collagen type were selected. According to online databases, these peptides are present in human, bovine, and rat collagens. As a result, this method can be potentially applied to other species' tissues as well, such as human tissues, and provides a universal and simple method of quantifying collagen types. The applicability of this method for analyzing collagen types was demonstrated on rat tissues (skin, tendon, and aorta).

Keywords: Collagen / Collagen types / HPLC-MS/MS / Proteomics

Received: June 18, 2008; revised: August 7, 2008; accepted: August 7, 2008

DOI 10.1002/jssc.200800351

1 Introduction

Collagens (and proteins with collagen-like domains) form large superfamilies in various species, and the numbers of known family members are constantly increasing. They are a family of extracellular matrix proteins that play a dominant role in maintaining the structure of various tissues and also have many other important functions (for example adhesion, tissue remodeling). Collagens are the most abundant proteins in the human body, constituting approximately 30% of its protein mass. Vertebrates have at least 27 collagen types with 42 distinct polypeptide chains, more than 20 additional proteins with collagen-like domains and approximately 20 isoenzymes of various collagen-modifying enzymes [1, 2]. The most abundant collagens form extracellular fibrils or network-like structures, but the others fulfill a variety of biological functions. Fibril-forming collagens represent a set of at least nine different polypeptide chains which constitute the molecular species of type (I–III, V, XI, XXIV, and XXVII) collagens.

Collagen type I is normally a heterotrimeric molecule, composed of two α1(I) chains and one α2(I) chain [α1(I)]2α2(I). It has been shown that the more hydrophobic α2(I) chain plays a role in stabilizing this heterotrimeric type I collagen [3]. On the other hand, homotrimeric collagen type I, composed of three α1(I) chains [α1(I)]3, has been shown to occur at low levels in normal adult skin [4], during embryonic development [5], and during wound healing [6]. This homotrimeric type I collagen has also been shown to be associated with certain forms of Ehlers-Danlos syndrome [7–9] and osteogenesis imperfecta [10, 11]. It has been reported that the presence of the homotrimeric type I collagen isotype significantly weakens the aorta [12, 13].

Type II collagen is often found in cartilage and the vitreous humor in association with type XI collagen [14]. It has been shown that some collagen type II peptides can serve as potential biomarkers of the activity of matrix metalloproteinases [15]. This activity leads to the destruction of cartilage, and therefore plays an important role in the pathology of osteoarthritis.

Collagen type III is codistributed with type I in tissues and there is a lot of information indicating that their relative proportions change under conditions of inflammation, during wound healing [16, 17], and under some pathological conditions [4, 18].

Type IV collagen is only found in basement membranes, where it is the major structural component. Abnormalities in this basement membrane collagen structure and function are connected to both inherited and acquired diseases. Type V usually occurs as a minor component alongside type I collagen fibers [19].

From the medical point of view, the analysis of collagens is of great importance. Over 400 mutations in six different collagens cause a variety of human diseases that include osteogenesis imperfecta, chondrodyspla-
sias, some forms of osteoporosis, some forms of osteoarthritis, and the renal disease known as Alport syndrome [1]. Considering all of the above-mentioned examples, it is not surprising that there is a desire for a simple method which would allow us to determine and quantify various collagen types or collagen chains.

From the analytical point of view, there are several ways of quantifying collagen and/or collagen types or even particular collagen chains. The most common method of determining total collagen is based on the quantitation of hydroxyproline, which accounts for approximately 10% of the collagen molecule [20–22]. A method based on a spectrophotometric assay of collagen stained by a strong anionic dye (Sirius red F3BA) in picric acid solution is also frequently used [23–25].

Other alternatives include radioactive labeling of proline and enzyme immunoassays using antibodies specific to each collagen type. Although these methods are very specific, they are expensive and many sample handling steps are required [26].

Other options, which enable scientists to determine the molar ratios of particular collagen types, employ the separation of the peptide mixture produced by enzymatic digestion by various separation methods, such as CE, SDS-PAGE, or HPLC [27] and their detection by UV, MS or MS/MS [28].

In this study, we employed the HPLCMS/MS analysis of marker peptides in the peptide mixture produced by cyanogen bromide/trypsin digestion.

2 Materials and methods

2.1 Calibration standards

The peptides used for calibration (GSEGPQGVR, GDQGPVGGR, GGAGPGPEGGK, where P represents hydroxylated proline) were synthesized by Vidia (Jesenice u Prahy, Czech Republic). The HPLC purity of peptides was 95.19, 98.84, and 99.10% for GSEGPQGVR, GDQGPVGGR, and GGAGPGPEGGK, respectively. The water used in the experiments was MilliQ, ACN (HPLC gradient grade) was obtained from Merck (Darmstadt, Germany). Mercaptoethanol, cyanogen bromide (CNBr), formic acid, and trypsin (lot 51K72501) were obtained from Sigma (St. Louis, MO, USA).

2.2 Tissues and animals used

Two types of collagen samples were used: (a) commercially available collagens from Sigma – type I c.n. C9879 (isolated from bovine achilles tendon), type II c.n. C1188 (isolated from bovine tracheal cartilage), type III c.n. C4407 (isolated from human placenta), type IV c.n. C3657 (isolated from human placenta) and (b) skin, tail tendon, and aorta from 6-month old male Lewis strain rats (average weight approximately 450 g) bred in the Institute of Physiology, Academy of Sciences of the Czech Republic, v.v.i. (Prague, Czech Republic).

2.3 Sample preparation

All commercially available collagen types I–V, and the lyophilized tissue samples (rat aorta, skin, and tendon) were treated the same way: incubation (24 h) in 0.1 M ammonium bicarbonate, pH 7.8, containing 25% v/v 2-mercaptoethanol to reduce the oxidized methionyl residues. The concentrations were as follows: 1 mg/mL for commercially available collagens and 2 mg/mL for tissue samples. After removing the 2-mercaptoethanol by lyophilization, the samples were cleaved with CNBr in 70% v/v formic acid under nitrogen for 72 h. The samples were repeatedly dried (with a controlled warm air flow) and reconstituted in water (3 times). After removing the CNBr, the samples were dissolved in 0.05 M ammonium bicarbonate buffer, pH 7.8, to achieve a concentration of 2 mg/mL for all samples, and treated with trypsin (1:50 enzyme/substrate ratio) at 37 °C for 3 h.

2.4 Separation conditions

Chromatographic separation was carried out in a C12 RP Jupiter Proteo 90 A column, 250 × 2 mm, particle size: 4 μm, pore size: 9 nm (Phenomenex, Torrance, CA, USA). The HPLC apparatus used was a HP 1100 LC system (Agilent, Palo Alto, CA, USA) consisting of a degasser, a binary pump, an autosampler, a thermostated column compartment, and a diode array detector. It was coupled to an IT mass spectrometer (Agilent LC-MSD Trap XCT-Ultra).

For peptide identification and selection experiments, separation was achieved via a linear gradient between mobile phase A (water/formic acid, 100:0.03 v/v) and B (ACN/formic acid, 100:0.025 v/v). Separation was started by running the system isocratically for two minutes with 2% mobile phase B, followed by a gradient elution to 35% B at 40 min. Finally, the column was eluted with 100% B for 10 min. Equilibration before the next run was achieved by washing with 2% mobile phase B for 10 min.

For analyses of selected peptides (quantification), the gradient was changed as follows to achieve the separation in a shorter period of time: separation was started by running the system isocratically for two minutes with 3% mobile phase B, followed by a gradient elution to 12% B at 15 min. Finally, the column was eluted with 100% B for 10 min. Equilibration before the next run was achieved by washing with 3% mobile phase B for 10 min.

The flow-rate for both types of experiments was 0.25 mL/min, injection volume was 40 μL, the column temperature was held at 25 °C and UV absorbance detection was done at 214 nm.
API-ESI positive mode IT MS was used. Operating conditions: drying gas (N₂), 10 L/min; drying gas temperature, 350 °C; nebulizer pressure, 25 psi.

For peptide selection experiments, ions were observed over the mass range m/z 100–2200 (MS: standard mode, MS/MS: enhanced mode). Analysis was done in auto MS/MS mode (ten precursor ions, excluded after two spectra for 0.5 min). The analyses of calibration standards and tissue samples were performed in multiple reaction monitoring mode (precursor ions were selected with respect to the peptides studied, see Section 3). The fragmentation amplitude was set to 1.14 V.

Analysis of MS/MS data (peptide/protein identification and searching for possible post-translational modifications) was carried out using SpectrumMill software (v.3.02, Agilent). Searches were performed in the full protein databases SwissProt and NCBI nr. To identify peptides with multiple hydroxylations, searches were performed on the data extracted from these databases.

2.5 Quantification

The quantification was performed using calibration solutions of marker peptides. The concentrations of marker peptides in tissue samples obtained were then recalculated using molecular weights to determine the concentration of the whole protein.

3 Results

3.1 Marker peptide selection

Five collagen standards (types I–V) were analyzed. For each collagen type, at least one characteristic tryptic peptide was selected, which was unique to the particular collagen type. All of these marker peptides had to meet all of the following criteria at the same time:

(a) They had to be detected in all standard samples of a given collagen type
(b) They had to be part of the collagenous domain.
(c) They had to be tryptic peptides, i.e., peptides obtained by cleavage with trypsin at the specific sites.
(d) Their peak area in HPLC-MS/MS chromatograms of the standard samples had to be reproducible.
(e) They had to be unique to a given collagen type, i.e., the peptide could not be part of any other protein structure in the SwissProt/NCBI nr databases.
(f) Their abundances had to be the same in all samples, i.e., one peptide per whole protein.
(g) They had to be common to at least three different species (bovine, human, and rat).

The retention times, masses of both parent and product ions for type I–V collagen marker peptides and their localization in the context of the primary structure of collagen are summarized in Table 1.

3.2 Sample analyses

Three types of tissue were analyzed (skin, tendon, and aorta) for their content of collagen types I–V. Only collagen types I and III were detected in these tissues, the other types were below the LOD. Furthermore, the ratios of α1(I)/α2(I) and those of collagen type I/III were determined.

The comparison of the molar ratios obtained showed no significant differences between the results obtained by the MS detection of precursor ions and MS/MS detection of product ions; however, MS/MS was used for quan-

<table>
<thead>
<tr>
<th>Peptide a)</th>
<th>Localization b)</th>
<th>Retention time (min) c)</th>
<th>Charge state</th>
<th>Precursor MH⁺</th>
<th>Precursor mass</th>
<th>Product ion mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collagen type I, α1 GSEGPQGVR</td>
<td>362–370</td>
<td>12.3</td>
<td>2⁺</td>
<td>886.3</td>
<td>443.7</td>
<td>613.2</td>
</tr>
<tr>
<td>Collagen type I, α2 GDQGPVGR</td>
<td>823–830</td>
<td>11.4</td>
<td>2⁺</td>
<td>785.3</td>
<td>393.2</td>
<td>485.2</td>
</tr>
<tr>
<td>Collagen type II, α1 TGPAGAAGAR</td>
<td>335–344</td>
<td>10.6</td>
<td>2⁺</td>
<td>828.3</td>
<td>414.7</td>
<td>335.5</td>
</tr>
<tr>
<td>Collagen type III, α1 GGAGPPEGGK</td>
<td>696–707</td>
<td>11.2</td>
<td>2⁺</td>
<td>996.4</td>
<td>498.7</td>
<td>377.6</td>
</tr>
<tr>
<td>Collagen type IV, α1 GPPGGVGFPGR</td>
<td>586–597</td>
<td>19.1</td>
<td>2⁺</td>
<td>1116.5</td>
<td>558.8</td>
<td>962.5</td>
</tr>
<tr>
<td>Collagen type V, α1 GPPCPAGPGER</td>
<td>1393–1403</td>
<td>13.3</td>
<td>2⁺</td>
<td>991.5</td>
<td>504.2</td>
<td>853.4</td>
</tr>
</tbody>
</table>

a) P, hydroxylated proline.
b) The numbering is done with respect to the human collagens with SwissProt accession numbers P02452, P08123, P02458, P02461, P02462, and P20908.
c) Gradient for peptide identification and selection experiments.
titration purposes due to its higher sensitivity and selectivity.

The calibration mixtures of the GSEGPQGVR, GDQGPVGR, and GGAGP peptides were analyzed, giving the following equations and correlation coefficients:

\[
\text{GSEGPQGVR}: A = 5.2078E+13 \times C + 8943897.16;
\]
\[
R^2 = 0.9992
\]
\[
\text{GDQGPVGR}: A = 8.0658E+13 \times C + 16216863.59;
\]
\[
R^2 = 0.9913
\]
\[
\text{GGAGPPGPEGGK}: A = 3.3781E+13 \times C + 1323897.96;
\]
\[
R^2 = 0.9997
\]

where \( A \) is the peak area of the product ion corresponding to a given peptide, and \( C \) is the concentration of peptide in g/mL. Each calibration consisted of eight different concentrations ranging from 1.01E-07 to 1.51E-05 g/mL, for GSEGPQGVR, GDQGPVGR, and GGAGPPGPEGGK, respectively.

The LOD and LOQ for the \( \alpha_1 \) and \( \alpha_2 \) chains of collagen type I and those for the \( \alpha_1 \) chain of collagen type III, which were determined based on the above-mentioned calibration data, are summarized in Table 2.

Two types of collagen (type I and III) were detected and quantified in all tissue samples, except in tail tendon samples, where collagen type III was detected, but could not be quantified due to its very low content, which was below the quantitation limit. Both the \( \alpha_1 \) and \( \alpha_2 \) chains of collagen type I were detected in all samples. The results are summarized in Table 3.

As far as the degree of hydroxylation of proline and lysine is concerned, only one form of peptide was found in both the collagen standard and the sample, unless stated otherwise. With collagen III, three forms of the GGAGPGPEGGK peptide were found in the collagen III standard samples: GGAGPGPEGGK, GGAGPGPEGGK, and GGAGPGPEGGK. However, the last two (m/z 506.7) were only found in standard samples of purchased collagen III. Furthermore, they were present in small quantities (~10% of the total peak area) compared to the GGAGPGPEGGK peptide. In biological samples, neither of these forms was detected.

<table>
<thead>
<tr>
<th>Collagen chain</th>
<th>LOD (g/mL)</th>
<th>LOQ (g/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 ), Type I</td>
<td>( 2.15 \times 10^{-6} )</td>
<td>( 7.16 \times 10^{-6} )</td>
</tr>
<tr>
<td>( \alpha_2 ), Type I</td>
<td>( 5.35 \times 10^{-7} )</td>
<td>( 1.78 \times 10^{-6} )</td>
</tr>
<tr>
<td>( \alpha_1 ), Type III</td>
<td>( 1.10 \times 10^{-6} )</td>
<td>( 3.67 \times 10^{-6} )</td>
</tr>
</tbody>
</table>

Table 2. LOD and LOQ for \( \alpha_1(I) \), \( \alpha_2(I) \), and \( \alpha_1(III) \) collagen chains

HPLC-MS/MS is a powerful tool not only for the identification, but also quantification of peptides and proteins. While this method has been previously employed for the identification of collagen types I and II [28], the quantification of collagen types and/or their ratios in tissues using HPLC-MS/MS is very rare. Instead, the quantification of collagen types is usually carried out by amino acid analyses. To date, the use of HPLC-MS/MS in this area was rather limited to the quantification of selected peptides of biological importance such as the quantification of collagen type II peptides [15]. In our work, we have extended the use of HPLC-MS/MS to quantify other collagen types using marker peptides.

The proposed method has some remarkable advantages over the methods that are generally used for the identification and/or quantification of collagens. For example, the main disadvantage of the widely used hydroxyproline and spectrophotometric assays is that they cannot determine particular collagen types. On the other hand, radioactive labeling and enzyme immunoassays using antibodies specific to each collagen type, though precise and selective, require very sophisticated (and expensive) equipment and therefore are not ordinarily available to everybody.

The method developed here represents a simple way of determining and quantifying the various collagen types in tissues using LC-MS/MS. Therefore, it combines the advantages of the above-mentioned methods and leaves out the problems associated with them.

4 Discussion
It was found that using the CNBr/trypsin digestion of collagen types I–V leads to peptide mixtures containing, among others, peptides characteristic for a given collagen type. Moreover, the versatility of the proposed method lies in the fact that each of these peptides is common to at least three species: human, bovine, and rat.

In our previous study, we showed that post-translational modifications could cause a problem due to resistance to enzymatic cleavage at a modified location, resulting in a lower yield of enzymatic digestion [29]. However, the search for 11 of the most common noncrosslinked post-translational modifications: imidazolone A (Δm/z: +144.04) and B (Δm/z: +142.03), Nε-(carboxymethyl)lysine (CML, Δm/z: +58.01), Nε-(carboxymethylhydroxy)lysine (CMhL, Δm/z: +74.00), Nε-(carboxyethyl)-lysine (CEL, Δm/z: +72.02), pyrraline (Δm/z: +108.02), 1-alkyl-2-formyl-3,4-glycosyl-pyrrole (AFGP, Δm/z: +270.07), Nδ-(5-hydroxy-4,6-dimethylpyrimidine-2-yl)-ornithine (argpyrimidine, Δm/z: +80.03), Nδ-(4-oxo-5-dihydroimidazol-2-yl)-ornithine or 1-(4-amino-4-carboxybutyl)-2-imino-5-oxo-imidazolodione (αNFC-1, Δm/z: +39.99), and Nδ-(5-methyl-4-oxo-5-hydroimidazol-2-yl)-ornithine or Nε-(4-methyl-5-oxo-4-

hydroimidazol-2-yl)-ornithine or 2-iminoimidazolidinedione (αNFC-1, Δm/z: +54.01) revealed that these peptides were either not modified in any of the samples measured or were modified to such a small extent that the modifications could not be detected.

The usability of the proposed method was demonstrated by determining the molar ratios of [α1(I)]2α2(I)/[α1(I)]3 and type I/III in real tissues (Table 3). The molar ratio of [α1(I)]2α2(I)/[α1(I)]3 was found to be 61:39 for aorta tissue, while for the skin and tendon samples, these ratios were 86:14 and 91:9, respectively. The purchased collagen type I (isolated from bovine achilles tendon) was found to be predominantly composed of heterotrimeric molecules ([α1(I)]/[α2(I)] = 2.11).

As has been shown previously by others, a higher percentage of the homotrimer results in a weakening of the tissues [12, 13]. Not surprisingly, the lowest amount of homotrimer was found in both achilles and tail tendon samples, which could reflect the fact that they are among the tissue types under the most strain, hence requiring
great tensile strength and, therefore, the lowest amount of homotrimer. On the other hand, the aorta sample, with the greatest amount of homotrimer, is the least strained of the three tissues studied, so the higher amount of homotrimer was to be expected.

The proportions of collagen types expressed as type I/III molar ratios for aorta and skin samples were found to be 69:31 and 86:14, respectively. While the values found for skin samples are in good agreement with older data previously found for human and rat skin in the literature [30, 31], they differ from the newer ones obtained by SDS-PAGE [32]. Considering the higher specificity and selectivity of the MS/MS technique, we believe that these recent results are more reliable than the previously published ones.

In tendon samples, collagen type III was detected but its content was below the quantitation limit, thus confirming other authors’ findings published for achilles tendon tissue [33, 34].

The analyses of collagen types II, IV, and V standard samples showed the possibility of quantitating these types as well. However, due to their lower occurrence in tissues, samples will need to be preconcentrated before they are analyzed.

The authors have declared no conflicts of interest.

This work was supported by the Grant Agency of the Czech Republic, grants nos. 203/06/1041, 203/08/1428, the Center for Cardiovascular Research IM0510, and by Research Project AV0Z50110509.

5 References