



HAL
open science

The chiral proteomic analysis applied to aging collagens by LC-MS: Amino acid racemization, post-translational modifications, and sequence degradations during the aging process

Marine Morvan, Ivan Mikšík

► **To cite this version:**

Marine Morvan, Ivan Mikšík. The chiral proteomic analysis applied to aging collagens by LC-MS: Amino acid racemization, post-translational modifications, and sequence degradations during the aging process. *Analytica Chimica Acta*, 2023, 1262, pp.341260. 10.1016/j.aca.2023.341260 . hal-04314111

HAL Id: hal-04314111

<https://hal.science/hal-04314111>

Submitted on 29 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

The chiral proteomic analysis applied to aging collagens by LC-MS: amino acid racemization, post-translational modifications, and sequence degradations during the aging process

Marine Morvan^{1,2,*}, Ivan Mikšík²

¹ Institute of Physiology of the Czech Academy of Sciences, Vídeňská 1083, 142 20 Prague, Czech Republic

² Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, Studentská 573, 532 10 Pardubice, Czech Republic

* corresponding author: marine.morvan@fgu.cas.cz (M. M.)

Citation: Morvan, M.; Mikšík, I., The chiral proteomic analysis applied to aging collagens by LC-MS: amino acid racemization, post-translational modifications, and sequence degradations during the aging process, *Analytica Chimica Acta* **2023**, *1262*, 341260. <https://doi.org/10.1016/j.aca.2023.341260>

Abstract: Collagen is the most abundant protein in the animal and human bodies, and it is not exempt from this aging phenomenon. Some age-related changes may appear on collagen sequences, such as increased surface hydrophobicity, the appearance of post-translational modifications, and amino acids racemization. This study has shown that the protein hydrolysis under deuterium conditions is privileged to limit the natural racemization during the hydrolysis. Indeed, under the deuterium condition, the homochirality of recent collagens is preserved whose amino acids are found in their L-form. However, in aging collagen, a natural amino acid racemization was observed. These results confirmed that the % D-amino acids are progressive according to age. The collagen sequence is degraded over time, and a fifth of the sequence information is lost during aging. Post-translational modifications (PTMs) in aging collagens can be a hypothesis to explain the modification of the hydrophobicity of the protein with the decrease of hydrophilic groups and the increase of hydrophobic groups. Finally, the exact positions of D-amino acids and PTMs have been correlated and elucidated.

Keywords: aging, chiral separation, collagen, D-amino acids, liquid chromatography, mass spectrometry.

1. Introduction

Aging is a natural and uncontrolled phenomenon that the animal body, including human, undergoes. *In vivo*, molecules are not exempt from this aging process. Indeed, some age-related protein changes may appear like a loss of proteolytic capacity [1]–[3], increased surface hydrophobicity [1], appearance of post-translational modifications such as oxidation [1], [4], [5], phosphorylation [4], methylation [4], [6], deamination [5], [7], and acylation (particularly acetylation [4], carbonylation [5], carboxymethylation [5]) [1], and racemization of amino acids [8]–[13].

Recent animal and human proteins are found in their L-amino acid forms. However, during the natural aging process, an amino acid racemization can take place in proteins *via* amino acids racemases (enzymatic process) or/and *via* a succinimidyl intermediate (non-enzymatic process) [14]. This amino acid racemization affects the three-dimensional protein conformation and can induce some aggregation [15]–[18], disorder [16], malfunctions [16], misfolding [16], and toxicity [16]. In addition, the amount of D-amino acids in proteins is progressive according to age. In 1975, D-Asp enrichment in human tooth enamel proteins *via* natural racemization was measured as a content of approximately 0,1 % per year [19]. Furthermore, some D-amino acids were related to aging human proteins and can be linked to aging diseases. It is the case of D-Asp in elastin linked to arteriosclerosis [20], [21], D-Asp in β -amyloid linked to Alzheimer's disease [22], D-Asp, D-Asn, D-Ser, and D-Thr in α -crystallin linked to cataract [9]. The complete list and the exact position of these D-amino acids in till now described protein sequences are recently summarized [14].

D-amino acids are also found in aging and/or ancient proteins and can be used for the age-estimation. Indeed, different studies used the quantification of D-Asp from bone [23], lens [24], and tooth [25], [26] for age-estimation, using a HCl/H₂O protein hydrolysis. However, in a hydrogenated environment, a natural racemization of L-amino acids to its enantiomer can take place. Different kinetics were observed according to the nature of amino acids [27], [28]. To prevent this amino acids racemization during the acidic hydrolysis, a DCl/D₂O condition is privileged [14], [24], [29], [30]. Indeed, in a deuterium environment, the hydrogen on the alpha carbon was exchanged with a deuterium atom and decreases considerably the racemization. Yasunaga *et al.* compared the racemization of D-Asp for age-estimation under HCl/H₂O and DCl/D₂O hydrolysis conditions. Errors in age estimations were decreased by at least one-half compared with that of the HCl/H₂O hydrolysis method [24].

Recently, chiral chromatographic and electrophoretic separation methods for the enantioseparation of (un-)derivatized D- and L-amino acids were described [14]. Mass spectrometry optimization can also enhance the discrimination of derivatized D- and L-amino acids, specifically using ion-mobility mass spectrometry [31], [32].

Collagen is the most abundant protein in animal bodies, including humans, and is the major structural component of the extracellular matrix present in tissues and organs [33], [34]. To date, 28 different types of collagens were described [33]. However, the most common of them is collagen type I which is found in 90 % [34]. Its three-dimensional structure is a triple-helix form made up of polypeptide chains α 1 and 2 (COL1A1 and COL1A2). The appearance of D-amino acids in the collagen protein reverses the rotation of the triple helix and decreases the helix-helix interactions until the overall destabilization of the three-dimensional structure [35], [36]. However, D-amino acids can also be involved in the cross-linking of assembly molecular material. Indeed, in aging collagen, the presence of D-Ala and D-Glu destabilized the structure while D-Lys stabilized it [2].

For this work, the amino acids racemization in collagen samples (type I alpha 1 and 2) from different organisms (bovine and rat) and at different ages (recent, 4-months and 4-years) were studied. Artificial aging was applied to standard bovine collagens to mimic natural aging. A comparison of HCl/H₂O and DCl/D₂O effects on collagen hydrolysis was studied. The amount of D-amino acids in bovine collagen after artificial aging was compared to that obtained in rat collagen all along the natural aging at the same age. After enzymatic treatment, arising peptides at different ages were compared. Common peptides between recent and aging collagen were identified as peptides from enzymatic digestions, and non-common peptides come from aging degradations. The evolution of post-translational modifications on the collagen sequences was also studied according to age. Finally, the exact position of D-amino acids and amino acids modified by post-translational modifications were elucidated.

2. Materials and methods

2.1. Chemicals

Acetic acid (AcOH, $\geq 99,8\%$), acetic acid-d₄ (DOAc, $\geq 99,9$ atom %D), acetonitrile (ACN, $\geq 99,9\%$), ammonium hydrogen carbonate (NH₄HCO₃, $\geq 99\%$), chloroform (99,8%), α -chymotrypsin from bovine pancreas, type II (≥ 40 units/mg protein), collagen from bovine achilles tendon, type I (bovine collagens), deuterium chloride solution (DCl, $\geq 99,9$ atom %D), deuterium oxide (D₂O, 99.9 atom %D), dimethyl sulfoxide (DMSO, $\geq 99,9\%$), ethyl acetate (EtOAc, $\geq 99,7\%$), formic acid (FA, $\geq 98\%$), heptafluorobutyric acid (HFBA, $\geq 99,5\%$), methanol (MeOH), N α -(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA, Marfey's reagent, $\geq 99\%$), pepsin from porcine stomach mucosa ($\geq 2,500$ units/mg protein (E1%/280)), phosphate buffer solution (1M, pH 7,4), proteinase K from *Tritirachium album* (≥ 30 units/mg protein), sodium bicarbonate (NaHCO₃, $\geq 99,5\%$), tris(hydroxymethyl)aminomethane (Tris, $\geq 99,8\%$), trypsin from porcine pancreas (13,000-20,000 BAEE units/mg protein) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Acetone ($\geq 99,5\%$) and sodium chloride (NaCl) were from Penta (Chrudim, Czech

Republic). Ultrapure water was obtained using a Milli-Q system from Millipore (Bedford, MA, USA)

Hereditary hypertriglyceridemic (HTG) rats were originally selected from Wistar rats as previously described [37]. HTG rats aged 4-months (bred in the Institute of Physiology, Academy of Sciences of the Czech Republic, Prague) were studied. Age-matched normotensive Wistar-derived Lewis rats were used as controls. Rats were housed under controlled conditions (temperature 23 ± 1 °C, 12 h : 12 h light-dark cycle). They were fed a standard rat chow (Velaz, ST-1) containing 0.4% sodium chloride. Water and food were available *ad libitum*.

Amino acid: D- α -alanine (D-Ala, $\geq 98\%$), D-arginine (D-Arg, $\geq 98\%$), L-arginine (L-Arg, $\geq 98\%$), D-aspartic acid (D-Asp, 99%), D-cysteine hydrochloride monohydrate (D-Cys, $\geq 98\%$), L-cysteine (L-Cys, 97%), D-glutamic acid (D-Glu, $\geq 99\%$), glycine (Gly, $\geq 99\%$), D-histidine monohydrochloride monohydrate (D-His, $\geq 98\%$), DL-5-hydroxylysine hydrochloride (DL-Hyl, $\geq 98\%$), trans-4-hydroxy-D-proline (D-Hyp, 97%), DL-isoleucine (DL-Ile, 99%), D-leucine (D-Leu, 99%), D-lysine (D-Lys, $\geq 98\%$), DL-methionine sulfoxide (DL-Met sulfoxide, $\geq 98.5\%$), L-methionine sulfoxide (L-Met sulfoxide, $\geq 98\%$), D-phenylalanine (D-Phe, $\geq 98\%$), D-proline (D-Pro, $\geq 99\%$), L-proline (L-Pro, $\geq 99\%$), DL-serine (DL-Ser, $\geq 98\%$), D-serine (D-Ser, $\geq 98\%$), D-threonine (D-Thr, $\geq 98\%$), L-threonine (L-Thr, $\geq 98\%$), D-tryptophan (D-Trp, $\geq 98\%$), L-tryptophan (L-Trp, $\geq 98\%$), D-tyrosine (D-Tyr, 99%), L-tyrosine (L-Tyr, $\geq 98\%$), and D-valine (D-Val, $\geq 98\%$) were obtained from Sigma-Aldrich (St. Louis, MO, USA). DL- α -alanine (DL-Ala), L-glutamic acid (L-Glu), DL-leucine (DL-Leu), DL-methionine (DL-Met), and DL- β -phenyl- α -alanine (DL-Phe) were obtained from Reachem (Mississauga, Canada). L-aspartic acid (L-Asp, 98%), L-histidine (L-His, 98%), L-methionine (L-Met, 98%) and L-valine (L-Val, 98%) were obtained from Roana (Budapest, Hungary). L-isoleucine (L-Ile, *allo*-isoleucine free), L-lysine monohydrochloride (L-Lys) were obtained from Calbiochem (San Diego, USA). L-hydroxyproline NH₂.CH₂.CH(OH).CH₂ (L-Hyp, $\geq 98.5\%$) was obtained from BDH (BDH Chemicals Ltd., Poole, UK).

2.2. Sample preparations

2.2.1. Standard and proteinogenic amino acids preparation

Standard amino acids preparation method: standard amino acids were dissolved in D₂O at 50 mM.

Amino acids derivatization method: To standard and proteinogenic amino acid solutions, 100 μ L of NaHCO₃ (1 M) in D₂O was added, and 200 μ L of 38.7 mM N α -(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA) in acetone. Solutions were vortexed and incubated at 40°C for 60 min. Reactions were quenched by the addition of 50 μ L of DCI (6 M) in D₂O. Compounds of interest are extracted by ethyl acetate (0.5 mL). A saturated sodium chloride solution (300 μ L) was added in the aqueous phase to increase the difference in density between the aqueous and organic phases and facilitate the extraction. Undesired hydrogen fluoride formed during the derivatization reaction stays in the aqueous phase and is not injected to limit the HPLC and LC column degradations.

From extracted organic phases, 20 μL of these solutions in 20 μL of DMSO were analyzed by HPLC at 340 nm or LC-MS.

2.2.2. Recent and aging collagens preparation

2.2.2.1. Artificial aging of bovine collagens

Bovine collagen samples ($n=6$) in phosphate buffer (100 mM, pH 7,4) at $2 \text{ mg}\cdot\text{mL}^{-1}$ in glass tubes, were placed outside under natural meteorological conditions for 4-months and 4-years.

2.2.2.2. Extraction of natural aging HTG rat collagens

Tendons from the tails of HTG rats ($n=2$) were scraped with a scalpel, then rinsed three times with 1M NaCl solution, each for 24h at 4°C . Extracted collagens were rinsed with ultrapure water and lyophilized.

2.2.2.3. Enzymatic digestion

Recent and aging bovine and rat collagens type I alpha 1 and 2 were digested successively by pepsin ($0.1 \text{ mg}\cdot\text{mL}^{-1}$ in 3 % AcOH, 50/1 sample/enzyme, w/w) and trypsin ($0.02 \text{ mg}\cdot\text{mL}^{-1}$ in 50 mM NH_4HCO_3 , 50/1 sample/enzyme, w/w). Collagen residues were lyophilized after both enzymatic treatments. Bovine and rat 4-months aging collagens produce an insoluble material after these treatments while recent and 4-years aging collagens do not. These insoluble parts were then digested by proteinase K ($0.05 \text{ mg}\cdot\text{mL}^{-1}$ in 10 mM Tris, 50/1 sample/enzyme, w/w), and resulting insoluble parts were digested by chymotrypsin ($0.04 \text{ mg}\cdot\text{mL}^{-1}$ in 50 mM NH_4HCO_3 , 50/1 sample/enzyme, w/w). The totality of bovine and rat 4-months aging collagens was soluble after chymotrypsin enzymatic treatment. Each enzymatic digestion was performed overnight at 37°C in the darkroom.

2.2.2.4. Hydrolysis

Proteins and peptides hydrolysis method: intact proteins and each peptide fraction were hydrolyzed in individual glass tubes placed into 10-mL Pierce Reacti-vials (Kimble Chase, Mexico) that had been purged with nitrogen thrice. After 18 h at 130°C in 400 μL 1:1 DCI/(2H4)acetic acid (DCI/DOAc) on the bottom of the vial, 200 μL of D_2O was added to the proteinogenic amino acid solution obtained in each tube. Then, these proteinogenic amino acids were derivatized following the amino acids derivatization method described in section 2.2.1.

2.3. HPLC parameters

2.3.1. HPLC for amino acids analysis

Analyses were performed in a HPLC system from Agilent (Agilent Technologies, Santa Clara, USA). Mobile phase A constituted to 0.1 % of FA in ultrapure water (v/v) and mobile phase B constituted to 0.1 % of FA in ACN (v/v). The chromatographic column was the Phenomenex Aeris

3.6 μm peptide XB-C18, 250 x 2.1 mm (Phenomenex Inc., Torrance, USA). The mobile phase operated in gradient mode, starting at 2 % mobile phase B and increasing for 30 min to reach 8 % B, 110 min to reach 40 %, and then increasing to 100 % B at 125 min, remaining for 20 min, and returning to 2 % B for 10 min maintaining constant until the end of the analysis. The flow rate was set at $0.250 \text{ mL}\cdot\text{min}^{-1}$, the column was held at ambient temperature (25°C), the injection volume was $5 \mu\text{L}$, and UV detection was performed at 340 nm.

The Agilent HPLC system was connected to a maXis quadrupole time-of-flight (Q-TOF) mass spectrometer with ultrahigh resolution (Bruker Daltonics, Bremen, Germany).

2.3.2. HPLC for peptides analysis

Analyses were performed on an HPLC system from Agilent (Agilent Technologies, Santa Clara, USA) consisting of a degasser, binary pump, autosampler, thermostatted column compartment and multi-wavelength detector. Mobile phase A constituted to 0.1 % of HFBA in ultrapure water (v/v) and mobile phase B constituted to 0.1 % of HFBA in ACN (v/v). Elution was achieved using a linear gradient when separation was started by running the system isocratically for two minutes with 2 % of mobile phase B, followed by a gradient elution to 35 % B at 40. min, Next gradient was 10 minutes to 100 % B. Finally, the column was eluted with 100 % B for 10 min. Equilibration before the next run was achieved by washing with buffer A for 10 min. The chromatographic column was the Jupiter 4 μm Proteo 90Å column (250 x 2 mm ID, Phenomenex Inc., Torrance, USA). The flow-rate was $0.250 \text{ mL}\cdot\text{min}^{-1}$, the column temperature was held at 25°C and UV absorbance detection was done at 214 nm, injection volume was $20 \mu\text{L}$.

Bovine and rat collagen peptides obtained after enzymatic treatments were separated and collected in several fractions, *e.g.* 31 fractions for collagen sample treated by pepsin and trypsin, 41 fractions after proteinase K treatment, and 37 fractions after chymotrypsin treatment. Chromatograms with fraction delimitations are presented in supplementary materials (Figures S1, S2, and S3).

2.4. NanoLC-MS parameters

The nanoHPLC apparatus used for protein digest analysis was a Proxeon Easy-nLC (Proxeon, Odense, Denmark). It was coupled to an ultrahigh resolution MaXis Q-TOF (quadrupole – time of flight) mass spectrometer (Bruker Daltonics, Bremen, Germany) by nanoelectrosprayer. The nLC-MS/MS instruments were controlled with the software packages HyStar 3.2 and micrOTOF-control 3.0. The data were collected and manipulated with the software packages ProteinScape 3.0 and DataAnalysis 4.0 (Bruker Daltonics).

Five microliters of the peptide mixture were injected into an NS-AC-12dp3-C18 Biosphere C18 column (particle size: $3 \mu\text{m}$, pore size: 12 nm, length: 200 mm, inner diameter: $75 \mu\text{m}$) with an NS-MP-10 Biosphere C18 precolumn (particle size: $5 \mu\text{m}$, pore size: 12 nm, length: 20 mm, inner diameter: $100 \mu\text{m}$), both manufactured by NanoSeparations (Nieuwkoop, Holland).

The separation of peptides was achieved via a linear gradient between mobile phase A (ultrapure water) and B (ACN), both containing 0.1 % FA (v/v). Separation was started by running the system with 5 % mobile phase B, followed by a gradient elution to 7 % B at 5 min, and 30 % B at 180 min. The next step was a gradient elution to 50 % B in 10 min and then a gradient to 100 % B in 10 min. Finally, the column was eluted with 100 % B for 20 min. Equilibration between the runs was achieved by washing the column with 5 % mobile phase B for 10 min. The flow rate was $0.200 \mu\text{L}\cdot\text{min}^{-1}$ and the column was held at ambient temperature (25 °C).

2.5. MS parameters

Online nano-electrospray ionization (easy nano-ESI) was used in positive mode. The ESI voltage was set to +4.5 kV, scan time: 3 Hz. Operating conditions: drying gas (N_2): $4 \text{ L}\cdot\text{min}^{-1}$; drying gas temperature: 180 °C; nebulizer pressure: 100 kPa. Experiments were performed by scanning from 50 to 2200 m/z. The reference ion used (internal mass lock, chip cube high mass reference (HP-1221), Agilent Technologies, Waldbronn, Germany) was a monocharged ion of $\text{C}_{24}\text{H}_{19}\text{F}_{36}\text{N}_3\text{O}_6\text{P}_3$ (m/z 1221.9906). Mass spectra corresponding to each signal from the total ion current chromatogram were averaged, enabling an accurate molecular mass determination. All HPLC-MS and nanoLC-MS analyses were done in duplicate.

2.6. Peptide identifications

Peptide data were processed using ProteinScape software (Bruker Daltonics, Bremen, Germany). Collagen type I alpha 1 and 2 chains were identified by correlating tandem mass spectra to the NCBI, IPI, and SwissProt databases, using the MASCOT search engine (<http://www.matrixscience.com>). The taxonomy was limited to *Bos taurus* and *Rattus norvegicus* due to the provenance of the samples. SemiTrypsin was chosen as the enzyme parameter. Three missed cleavages were allowed, and an initial peptide mass tolerance of ± 15.0 ppm was used for MS and ± 0.03 Da for MS/MS analysis. Arginine, asparagine, and glutamine were assumed to be deamidated, methionine to be dioxidized, lysine, methionine, and proline to be oxidized, lysine, serine and threonine to be formylated, lysine to be acetylated, carbamylated, carboxylated and carboxymethylated, arginine, aspartic acid, cysteine, histidine, lysine, serine, threonine, and tyrosine were allowed to be phosphorylated. All these possible modifications were set to be variable. Monoisotopic peptide charge was set at 1+, 2+, and 3+. The Peptide Decoy option was selected during the data search process to remove false-positive results. Only significant hits (accepted as identified on ProteinScape) were selected.

3. Results and discussion

3.1. Effect of hydrolysis conditions

Figure 1 shows the amino acid racemization of recent bovine collagens under HCl/H₂O (Figure 1A) and DCl/D₂O (Figure 1B) hydrolysis conditions. Amino acid racemization takes place only under HCl/H₂O hydrolysis conditions for all amino acids except Ser and Thr which were not

racemized. Methionine, phenylalanine, tryptophan, and tyrosine were not detected. For this work, a DCI/D₂O hydrolysis is privileged to limit the racemization during the hydrolysis and not distort the results when determining the % D-amino acids.

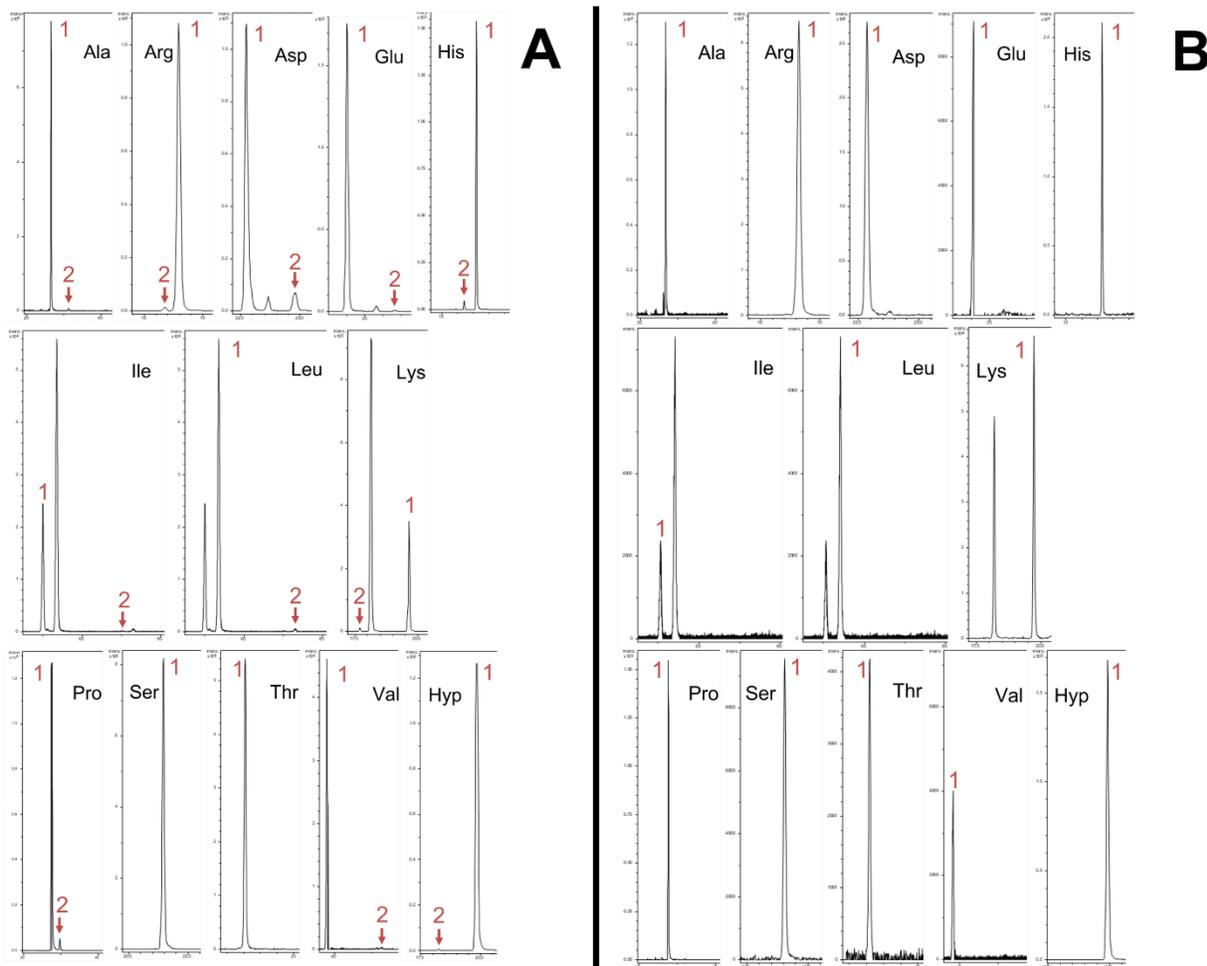


Figure 1: Effect of hydrolysis conditions on proteinogenic amino acid racemization from standard bovine collagens. Hydrolysis conditions: HCl/H₂O (A) and DCI/D₂O (B). Peak 1 identifies amino acids in their L-form and Peak 2 in their D-form. Exact masses of derivatized amino acids: 342,104420 (Ala), 427,168420 (Arg), 386,094250 (Asp), 400,109900 (Glu), 408,126220 (His), 384,151370 (Ile/Leu), 651,211740 (Lys), 368,120070 (Pro), 358,099340 (Ser), 372,114990 (Thr), 370,135720 (Val) and 384,114990 (Hyp).

3.2. D-amino acids in aging collagens

This chiral analysis method, described in section 2.3.1 and newly developed for this work using standard amino acids, was then used to confirm the presence of D-amino acids in bovine collagen after *in vitro* artificial aging, at different ages. In addition, this analytical method was also used for the detection of D-amino acids in rat collagen with *in vivo* natural aging. Preliminary study was

performed to confirm the presence of amino acid only in their L-form in recent bovine collagen (Table S5). During the aging process, the artificial amino acid racemization, applied to bovine aging collagens, mimics the natural amino acid racemization. Most of the different amino acids were racemized over time. It is the case of D-Arg, D-Asp, D-Leu, D-Pro, D-Ser, D-Thr, and D-Val most frequently racemized in 4-years than in 4-months collagen samples. In addition, the percentage of Arg, His, Thr, and Val in their D-form was higher in the older sample (Figures 2 and 3, Figures S4 and S5, Table S1 – S5).

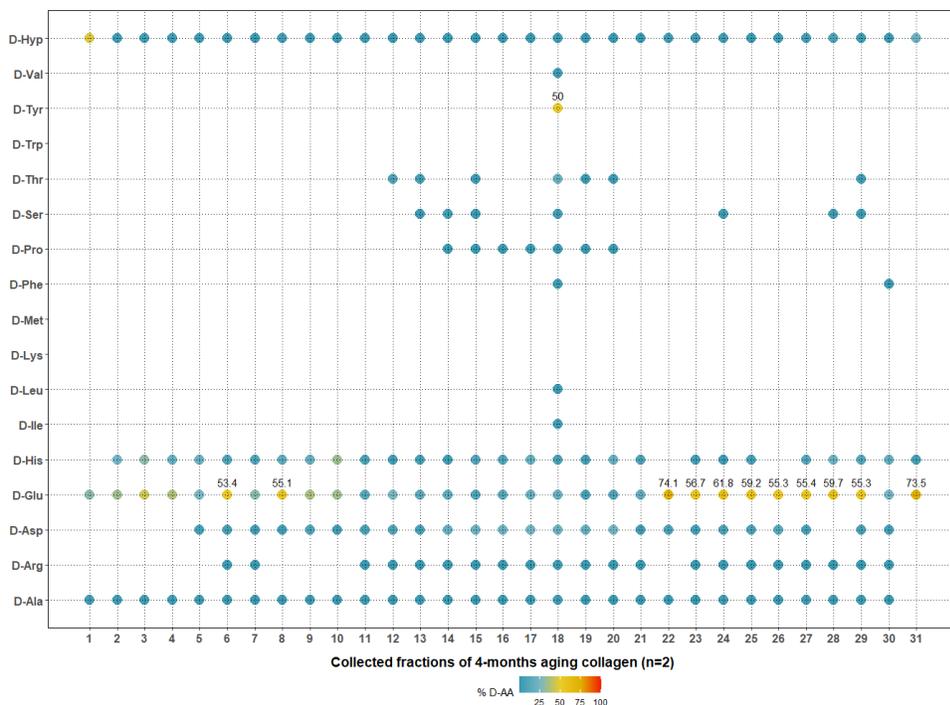


Figure 2: Representation of % D-amino acids in 4-months aging bovine collagen fractions (n=2) after pepsin and trypsin enzymatic treatment. Percentages of D-amino acids superior to 50 % are indicated on the graph. All percentages of D-amino acids are summarized in Table S1. The total % RSD of %D-amino acids was 4.37 %.

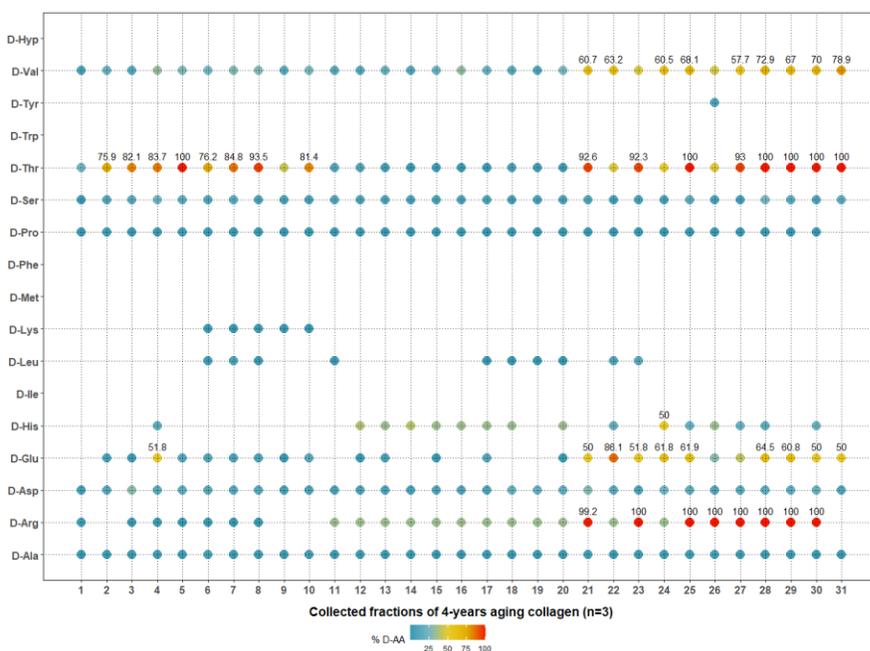


Figure 3: Representation of % D-amino acids in 4-years aging bovine collagen fractions (n=3) after pepsin and trypsin enzymatic treatment. Percentages of D-amino acids superior to 50 % are indicated on the graph. All percentages of D-amino acids are summarized in Table S2. The total % RSD of %D-amino acids was 10.97 %.

Most interesting amino acid fractions of bovine collagens were selected and compared to the same amino acid fractions of rat collagens at the same age. The selection of fractions was made based on the high percentage of amino acid racemization, and the diversity of the nature of amino acids racemized. The comparison of the percentage of D-amino acids is shown in Figures S6 and S7 and table S2 and S6. Into selected fractions of soluble part of collagen after pepsin and trypsin enzymatic treatment, some D-amino acids, *e.g.* D-Arg, D-Ile, D-Leu, D-Phe, D-Pro, and D-Tyr, were present in bovine collagen after artificial aging and not in rat collagen with natural aging. Other amino acids were less frequently present in rat collagen with natural aging, such as D-Asp, D-His, D-Ser, and D-Hyp. Conversely, D-Thr and D-Val were more frequently present in rat collagen with natural *in vivo* aging. Concerning the % D-amino acids, D-Glu, D-His, D-Thr, and D-Val were most racemized in their D-form in rat collagen with natural aging. The same trend was observed in selected fractions of the insoluble part of bovine and rat collagens after pepsin, trypsin, and proteinase K enzymatic treatment (Figure 4 and Table S3 and S7). Some D-amino acids, *e.g.* D-Arg, D-Lys, D-Phe, and D-Pro, were present in bovine collagen and not in rat collagen. D-Asp, D-His, D-Ser, and D-Hyp were less frequently present in rats. Conversely, D-Thr was more frequently present in rat collagen. Concerning the % D-amino acids, D-Glu and D-Thr were most racemized in their D-form in rat collagen with the natural aging, and D-Val was most racemized in bovine collagen after the artificial aging.

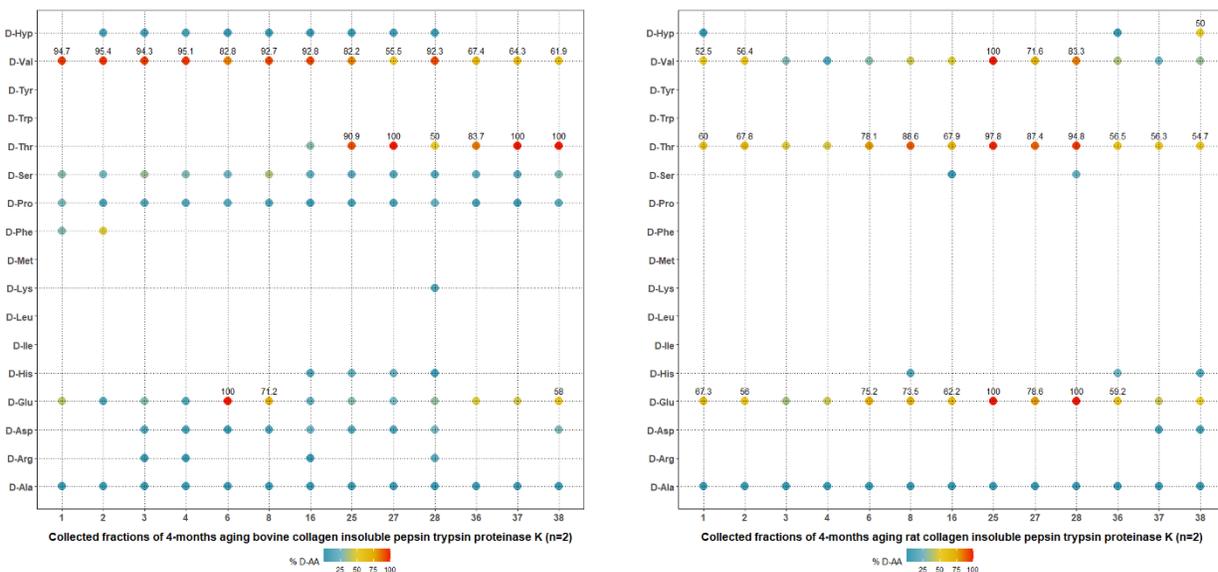


Figure 4: Representation of % D-amino acids in selected fractions of bovine and rat aging collagen after pepsin, trypsin, and proteinase K enzymatic treatment. Percentages of D-amino acids superior to 50 % are indicated on the graph. All percentages of D-amino acids are summarized in Table S3 and S7. The total % RSD of %D-amino acids from rat collagen was 7.07 %.

3.3. Peptide analysis

Peptides in each fraction from recent and aging collagens were analyzed by nanoLC-MS (sections 2.4 and 2.5) and identified (section 2.6)

3.3.1. Position of D-amino acids

The identification of peptides in each collected fraction (for the method see section 2.6) allowed us to determine their position on the bovine and rat collagen sequences (type I alpha 1 and 2). Regrouping the racemization rate of each amino acid in each fraction (section 3.2) and the peptide identification in these same fractions, the exact position of amino acids totally racemized in their D-form have been determined on the bovine and rat collagen sequences (Figure 5, 6, S8 and S9, Table 1 and S8).

In bovine collagens, 21 specific sites of Thr were identified as totally racemized in the D-form for 4-months aging sample and 18 specific sites for 4-years aging sample in the COL1A1 collagen sequence. Their exact positions have shown that 70 % of them were common at both ages. For COL1A2, 14 and 15 Thr specific sites were found in the 4-months and 4-years aging samples respectively, and 45 % were common at both ages. These specific sites of D-Thr totally racemized represent 48 % (4-months) and 41 % (4-years) of total Thr on the bovine COL1A1 sequence. For bovine COL1A2, these D-Thr specific sites represent 32 % (4-months) and 34 % (4-years) of total Thr on the protein sequence. D-Val was exclusively totally racemized in 4-months aging collagens and represent 19 % of total Val on bovine COL1A1 and 4 % on bovine COL1A2 sequences. D-

Arg was entirely racemized in 4-years aging collagens and represented 75 % of total Arg on bovine COL1A1 and 63 % on bovine COL1A2 sequences.

For amino acids totally racemized in fractions of 4-month aging rat collagens selected according to most promising bovine fractions at the same age, 12 specific sites for D-Glu (17 %) and 6 for D-Val (14 %) were identified in COL1A1. In COL1A2, 10 specific sites for D-Glu (16 %) and 6 for D-Val (12 %) were identified.

Table 1: Exact position of D-amino acids totally racemized in their D-form on the bovine collagen sequences at different ages.

		COL1A1				COL1A2	
AA	Position	4-months	4-years	AA	Position	4-months	4-years
R	131		✓	R	31		✓
	186		✓		38		✓
	219		✓		41		✓
	239		✓		130		✓
	243		✓		154		✓
	252		✓		163		✓
	267		✓		178		✓
	303		✓		232		✓
	309		✓		280		✓
	311		✓		325		✓
	321		✓		340		✓
	360		✓		379		✓
	369		✓		397		✓
	414		✓		411		✓
	468		✓		421		✓
	471		✓		430		✓
	486		✓		438		✓
	510		✓		448		✓
	527		✓		474		✓
	563		✓		484		✓
	573		✓		571		✓
	597		✓		586		✓
	630		✓		607		✓
	675		✓		643		✓
	684		✓		661		✓
	696		✓		669		✓
	732		✓		673		✓
	744		✓		691		✓
	762		✓		706		✓
	795		✓		739		✓
	801		✓		775		✓
	864		✓		792		✓
	881		✓		828		✓

	917		✓		877		✓
	957		✓		904		✓
	966		✓		924		✓
	969		✓		946		✓
	993		✓		976		✓
	1013		✓		994		✓
	1025		✓		1003		✓
	1035		✓		1015		✓
	1065		✓		1051		✓
	1083		✓		1078		✓
	1092		✓		1116		✓
	1104		✓		1203		✓
	1140		✓		1216		✓
	1167		✓				
	1169		✓				
	1213		✓				
	1435		✓				
T	117	✓		T	29		✓
	327	✓	✓		168	✓	✓
	336	✓	✓		207		✓
	444		✓		211		✓
	453	✓	✓		337	✓	✓
	540	✓	✓		382	✓	✓
	552	✓	✓		424	✓	
	765	✓			561	✓	
	792	✓	✓		601	✓	✓
	873	✓	✓		658	✓	
	912	✓	✓		741	✓	✓
	947	✓	✓		754	✓	
	1038	✓			784	✓	✓
	1068	✓	✓		793	✓	
	1097		✓		832	✓	✓
	1170	✓	✓		852	✓	✓
	1232	✓	✓		858	✓	✓
	1233	✓	✓		1319		✓
	1366	✓			1325		✓
	1380	✓			1331		✓
	1430	✓	✓				
	1431	✓	✓				
	1433	✓	✓				
V	599	✓		V	156	✓	
	606	✓			157	✓	
	686	✓					
	921	✓					
	1056	✓					

1	MFSFVDLRLL	LLLAATALLT	HGOEEGQEEG	QEEDIPTVTC	VQNGRLYHDR	50	1	MFSFVDLRLL	LLLAATALLT	HGOEEGQEEG	QEEDIPTVTC	VQNGRLYHDR	50
51	DVWKPVPCQI	CVCNDGNVLC	DDVICDELKD	CPNAKVPTDE	CCPVCPEGQE	100	51	DVWKPVPCQI	CVCNDGNVLC	DDVICDELKD	CPNAKVPTDE	CCPVCPEGQE	100
101	SPTDQETTGV	EGPKGDT CPR	CPRGFAGPPG	RDGIPGQPLG	PGFPFGPPPP	150	101	SPTDQETTGV	EGPKGDT CPR	CPRGFAGPPG	RDGIPGQPLG	PGFPFGPPPP	150
151	GPPGLGGNFA	PQLSYGYDEK	STGISVPGFM	GPSGPRGLPG	PPGAPPGQGF	200	151	GPPGLGGNFA	PQLSYGYDEK	STGISVPGFM	GPSGPR GL EG	PPGAPPGQGF	200
201	QGPPGEGPEE	GASGPMGPRG	FPGPPKRNMD	DGEAGKPRRP	GERGPPGPGQ	250	201	QGPPGEGPEE	GASGPMGPRG	FPGPPKRNMD	DGEAGKPR R	GERGPPGPGQ	250
251	ARGLPGTAGL	PGMKGHRGFS	GLDGAAGDAG	FAGPKGKPGS	PGENGAFGQM	300	251	ARGLPGTAGL	PGMKGHRGFS	GLDGAAGDAG	FAGPKGKPGS	PGENGAFGQM	300
301	GPRGLPGERG	RPGAPGPAGA	RGNDGAT GAA	GPPGPT GPAC	FPGFPGAVGA	350	301	GPRGLPGERG	RPGAPGPAGA	RGNDGAT GAA	GPPGPT GPAC	FPGFPGAVGA	350
351	KGRGGPQGR	GSEGPQVVRG	EPGPPGPAGA	AGPAGNPGAD	GQPKAGKANG	400	351	KGRGGPQGR	GSEGPQVVRG	EPGPPGPAGA	AGPAGNPGAD	GQPKAGKANG	400
401	APGIAGAPGF	PGARGSPGQ	GSPGPPGPKG	NSGEPGAPGS	KGDYGAKEFP	450	401	APGIAGAPGF	PGARGSPGQ	GSPGPPGPKG	NSGEPGAPGS	KGDYGAKEFP	450
451	GPTCIQGGPPG	PAGEECKRCA	RCEPCACLEP	CPPECRCPCP	SRCFPCADCV	500	451	GPTCIQGGPPG	PAGEECKRCA	RCEPCACLEP	CPPECRCPCP	SRCFPCADCV	500
501	AGPKGPAGER	GAPCPAGPKG	SPGEARPCPE	AGLPCAKGLT	GSPGSPGPDG	550	501	AGPKGPAGER	GAPCPAGPKG	SPGEARPCPE	AGLPCAKGLT	GSPGSPGPDG	550
551	KTGPPGPAGQ	DGRGPPGPP	GARGQAGVMG	FPGPKAAGE	PGKAGERGVF	600	551	KTGPPGPAGQ	DGRGPPGPP	GARGQAGVMG	FPGPKAAGE	PGKAGERGVF	600
601	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQQPAGSPG	FQGLPGPAGP	650	601	GPPGAVGPAG	KDGEAGAQQP	PGPAGPAGER	GEQQPAGSPG	FQGLPGPAGP	650
651	PGEAGKPEQ	GVPDGLGAPG	PSGARGERGF	FERGV VOGPP	GPAGPRGANG	700	651	PGEAGKPEQ	GVPDGLGAPG	PSGAR R GERGF	FERGV VOGPP	GPAGPRGANG	700
701	APGNDGAKGD	AGAPYAPGSG	GAPGLQGMFG	FRGAAGT.PGP	KGDRGDAGPK	750	701	APGNDGAKGD	AGAPYAPGSG	GAPGLQGMFG	FRGAAGT.PGP	KGDRGDAGPK	750
751	GADGAPGKDG	VRGLTGP IGP	FGPAGAPGDK	GEAGSPGAPG	PTGARGAPGD	800	751	GADGAPGKDG	VRGLTGP IGP	FGPAGAPGDK	GEAGSPGAPG	PTGARGAPGD	800
801	RCEPCPPCPA	CFACPPGADG	QPCAKGEPGD	AGAKDCAFP	GPACAPGPPG	850	801	RCEPCPPCPA	CFACPPGADG	QPCAKGEPGD	AGAKDCAFP	GPACAPGPPG	850
851	PIGNVGAAPG	KGARGSAGPP	GATGFPPGAG	RVGPPGPGSN	AGPPGPPGPA	900	851	PIGNVGAAPG	KGARGSAGPP	GATGFPPGAG	RVGPPGPGSN	AGPPGPPGPA	900
901	GKEGSKGPRG	ETGPAGRPCE	VGGPPGPGPA	GEKGAPGADG	PAGAPCT PGP	950	901	GKEGSKGPRG	ETGPAGRPCE	VGGPPGPGPA	GEKGAPGADG	PAGAPCT PGP	950
951	QGIACQRCVV	GLPQRCGERG	FPGLPGPSGE	PGKQGFSCAS	GERGPPCPMG	1000	951	QGIACQRCVV	GLPQRC LE RG	FPGLPGPSGE	PGKQGFSCAS	GERGPPCPMG	1000
1001	PPGLAGPPGE	SGREGAPGAE	GSPGRDGSFG	AKGDRGET GP	AGPPGAPGAP	1050	1001	PPGLAGPPGE	SGREGAPGAE	GSPGRDGSFG	AKGDRGET GP	AGPPGAPGAP	1050
1051	GAPGPVGPAG	KSGDRGET GP	AGPAGPIGPV	GARGPAGPQG	PRGDKGET GE	1100	1051	GAPGPVGPAG	KSGDRGET GP	AGPAGPIGPV	GARGPAGPQG	PRGDKGET GE	1100
1101	QDNRG1KGHR	GISGLQGPPG	PPGSPGEQGP	SGASGPAGPR	GPPGSAGSPG	1150	1101	QDNRG1KGHR	GISGLQGPPG	PPGSPGEQGP	SGASGPAGPR	GPPGSAGSPG	1150
1151	KDGLNGLPGP	IGPPGPRC RT	GDAGPAGPPG	FPGPPGPPGP	PSGGYDLSFL	1200	1151	KDGLNGLPGP	IGPPGPRC RT	GDAGPAGPPG	FPGPPGPPGP	PSGGYDLSFL	1200
1201	PQPPQEKAKHD	GGRYRADD	NVVRDRDLEV	DTTLKSL SQQ	TENIRSPRGS	1250	1201	PQPPQEKAKHD	GGRYRADD	NVVRDRDLEV	DTTLKSL SQQ	TENIRSPRGS	1250
1251	RKNPARTCRD	LKMCHSDWKS	GEYWDPNQG	CNLDAIKVFC	NMETGETCVY	1300	1251	RKNPARTCRD	LKMCHSDWKS	GEYWDPNQG	CNLDAIKVFC	NMETGETCVY	1300
1301	PTQPSVAQKN	WYISKNPKEK	RHWYGESMT	GGFQFEYGGQ	GSDPADVAIQ	1350	1301	PTQPSVAQKN	WYISKNPKEK	RHWYGESMT	GGFQFEYGGQ	GSDPADVAIQ	1350
1351	LTFLRLMSTE	ASQNTYHCK	NSVAYMDOOT	GNLKKALLQ	GSNEIEIRAE	1400	1351	LTFLRLMSTE	ASQNTYHCK	NSVAYMDOOT	GNLKKALLQ	GSNEIEIRAE	1400
1401	GNSRFTYSVT	YDGTSHHTGA	WGKTVIEYK T	TKTSRL PIID	VAPLDVGAPD	1450	1401	GNSRFTYSVT	YDGTSHHTGA	WGKTVIEYK T	TKTSRL PIID	VAPLDVGAPD	1450
1451	QEPGFDVGPA	CFL				1463	1451	QEPGFDVGPA	CFL				1463

Figure 5: Exact position of D-amino acids totally racemized in their D-form on the bovine collagen COL1A1 sequence at 4-months (left) and 4-years (right).

1	MFSFVDLRLL	LLLAATALLT	HGQEEGQEEG	QEEDIPPVTC	VQNGRLYHDR	50	1	MFSFVDLRLL	LLLGATALLT	HGQEDIPEVS	CIHNGLRVFN	GETWKPDVCL	50
51	DVWKVPVPCQI	CVCDDNGNVLIC	DDVICDELKID	CNPNAKVPIDE	CCPCVPEGQE	100	51	ICICHNGTAV	CDGVLCKEDL	DCPNPQKREG	ECCFPCEEY	VSPDAEVIQV	100
101	SPTDQETTVG	EGPKGDTCTPR	GPRGFPAGPPG	RDGIPGQPLG	PCPPGPPGPP	150	101	EGPKGDGPGQ	GPRGFPVGGP	QDGIQGPGL	PGPPGPPGPP	GPPGLGGNFA	150
151	GPPCLGKNFA	PQLSYCYDEK	STGISVPGPM	GPSGPRGLPG	PPGAPGQGF	200	151	SQMSYGYDEK	SAGVSVPGPM	GPSGPRGLPG	PPGAPGQGF	QGGPGEPEP	200
201	QCPCPEPCEP	CASGPMCPFR	PPGPPKRNCD	DCEACKPCRP	CERGPCCPQC	250	201	GASGPMGPRG	PPGPPKRNCD	DCEAGKPCRP	GERGPPGQGF	ARGLPQTAGL	250
251	ARGI.PGTAGI	PGMKGHRGFS	GLDGAAGDAG	PAGPKGFRPGS	PGENGAPQM	300	251	PGMKGHRGFS	GLDGAAGDAG	PAGPKGFRPGS	PGENGAPQM	GPRGLPGERG	300
301	GPRGLPGERG	RPGAGPPAGA	RNGDGAATGAA	GPPGPTGPAG	PPGPPGAVGA	350	301	RPGPPGSAGA	RNGDGAAGAA	GPPGPTGPAG	PPGPPGAVGA	KGEAGFPQGAR	350
351	KGEGGPPQGR	GSEGPQCVRG	EPGPPFPAGA	AGPACNPGAD	QQFGAKGANG	400	351	GSEGPQVVRG	EPGPPFPAGA	AGPACNPGAD	GQFGAKGANG	APGIAGAPGF	400
401	APGIAGAPGF	PGARGPSGQ	GSPGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	450	401	PGARGPSGQ	GSPGAPGPKG	NSGEPGAPGN	KGDTGAKGEP	GPAGVQPPPG	450
451	GPTGIQGFPG	PAGEEGKRGGA	RGEPPGAGLP	GPPGERGGPG	SRGFPAGDGV	500	451	PAGEEGKRGGA	RGEPPGAGLP	GPPGERGGPG	SRGFPAGDGV	AGPKGPAGER	500
501	AGPKGPAGER	GAPGAPGPKG	SPGEAGRPEG	AGLPGAKGLT	GSPGSPGPDG	550	501	GSPGAPGPKG	SPGEAGRPEG	AGLPGAKGLT	GSPGSPGPDG	KTGPPGAPAG	550
551	KTCPPGPPAQ	DGRGPPGPPG	GARGQAGVMG	FPGPKGAAGE	PKKACERGVF	600	551	DGRGPPGAPG	GARGQAGVMG	FPGPKGTAGE	PKKACERGVF	GPPGAVGPPG	600
601	GPPGAVGPPG	KDGEACAQCP	PGPAGPAGER	GEQQPAGSPG	FQGLPGPACP	650	601	KDGEAGAQA	PGPAGPAGER	GEQQPAGSPG	FQGLPGPACP	PGEAGKPEGE	650
651	PGEAGKPEGE	GVPDGLGAPG	PSGARGRRGF	PFGRVQGGP	GPAGPRGANG	700	651	GVPDGLGAPG	PSGARGRRGF	PFGRVQGGP	GPAGPRGANG	APNGDGAAG	700
701	APNGDGAAG	AGAPYAPGQ	GAPGIQGMGP	FRGAAGI.PGP	KGDRGDAGPK	750	701	TGAPGAPGQ	GAPGLQGMGP	FRGAAGI.PGP	KGDRGDAGPK	GADGSPKDG	750
751	GADGSPKDG	VRGLTPIGP	PGPAGAPGK	GEAGPSGAPG	PTGARGAGPD	800	751	VRGLTPIGP	PGPAGAPGK	GEAGPSGAPG	PTGARGAGPD	RGEFPGPPGA	800
801	RCEFPGPPGA	CFACFPACADG	QPCAKCEPCD	ACAKDAGAPP	GFAGPACPPG	850	801	GFAGPPGADG	QPCAKCEPCD	TGVKGDAGPP	GFAGPACPPG	PIGNVGAAGP	850
851	PIGNVGAAGP	KGARGSACPP	GATGFPGAAG	RVGPPGPPSGN	AGPPGPPGPA	900	851	KGSRGAAGPP	GATGFPGAAG	RVGPPGPPSGN	AGPPGPPGPPV	GKEGKGGPRG	900
901	GKEGKGGPRG	ETGPACRPE	VGDPGPPGPA	GEKAPGADG	PAGAPCTPPI	950	901	ETGPACRPE	VGPPGPPGPA	GEKAPGADG	PAGSPGTPP	QGIAGQRGVV	950
951	QGIAGQRGVV	GLPQQRGERG	FPLGPPSGE	PGKQPSGAS	GERGPPGPMG	1000	951	GLPQQRGERG	FPLGPPSGE	PGKQPSGAS	GERGPPGPMG	PPGLAGPPGE	1000
1001	PPGLAGPPGE	SCRECAPCAE	GSPGRDGSFC	AKCDRETCPE	AGPPGAFGAP	1050	1001	SGREGSPGAE	GSPGRDGA	AKCDRETCPE	AGPPGAPGAP	GAPGPPGAPG	1050
1051	GAPGPPGAPG	KSGDRGCTGP	AGFAGPIGVP	GARGPAGPQC	PRDKKGTGE	1100	1051	KNGDRGETGP	AGPAGPIGPA	GARGPAGPQC	PRDKKGTGE	QDGRGKGGHR	1100
1101	QDGRGKGGHR	GFSGIQGGPP	PPGSPGRQGP	SGASGPAGPR	GPPSAGSPG	1150	1101	GFSGLQGGPP	SPGSPGEGQP	SGASGPAGPR	GPPSAGSPG	KDGLNGLP	1150
1151	KDGLNGLP	IGPPGPRGR	GDAGPAGPPG	PPGPPGPPGP	PSGGYD.SFI	1200	1151	IGPPGPRGR	GDSGPAGPPG	PPGPPGPPGP	PSGGYD.SFI	PQPPQEKSD	1200
1201	PQPPQEKSD	GCRYRADA	NVVRDRDLEV	DTLKLKLSQ	IENIRSPGCS	1250	1201	GCRYRADA	NVVRDRDLEV	DTLKLKLSQ	IENIRSPGCS	RKNPARTCRD	1250
1251	RKNPARTCRD	LKMCHSDWKS	GEYWDENQC	CNLDAIKVFC	NMETCETCVY	1300	1251	LKMCHSDWKS	GEYWDENQC	CNLDAIKVFC	NMETGTCTVF	PTQPSVPPGKN	1300
1301	PTQPSVPPGKN	WYISKNPKEK	RHWYGESMT	GGFQFEYGGQ	GSDPADVAIQ	1350	1301	WYISKNPKEK	RHWYGESMT	GGFQFEYGGQ	GSDPADVAIQ	LTFLRLMSTE	1350
1351	LTFLRLMSTE	ASQNTIYHCK	NSVAYMDQQT	GNLKKALLQ	GSNETETRAE	1400	1351	ASQNTIYHCK	NSVAYMDQQT	GNLKKALLQ	GSNETELRGE	GNSRFTYSTL	1400
1401	GNSRFTYSTL	YDGCSTHTGA	WGKTVEYKT	TKTSRLPIID	VAPLDVGA	1450	1401	YDGCSTHTGT	WGKTVEYKT	TKTSRLPIID	VAPLDVGA	QEFGMDIGPA	1450
1451	QEFGMDIGPA	CFL				1463	1451	CFV				1453	

Figure 6: Exact position of D-amino acids totally racemized in their D-form on the bovine collagen COL1A1 sequence at 4-months (left) and on the rat collagen COL1A1 sequence at 4-months (right).

3.3.2. Peptides from enzymatic digestion and aging degradation

After successively enzymatic digestions by pepsin and trypsin, all peptides and residues from recent and aging bovine collagens (COL1A1 and COL1A2) were soluble and identified by nanoLC-MS (section 2.4). Peptides from the insoluble part of collagen protein after pepsin and trypsin treatment at 4-months of aging were not added in the study. Arising peptides in aging soluble collagen samples common to arising peptides in recent collagen were identified as peptides from enzymatic digestions. Other peptides were identified as peptides from aging degradation (Figure 7). For aging bovine collagens type I alpha 1 and 2 (COL1A1 and COL1A2), the number of peptides from aging degradation is higher than arising peptides from enzymatic digestions. Indeed at 4-months, 46 % (COL1A1) and 48 % (COL1A2) of peptides come from aging degradations for both collagen types, and 55 % (COL1A1) and 46 % (COL1A2) at 4-years. The total of arising peptides from enzymatic digestion and aging degradation did not reach 100 % for both collagen chains and at both ages. Approximately, a fifth of the sequence information was lost during aging.

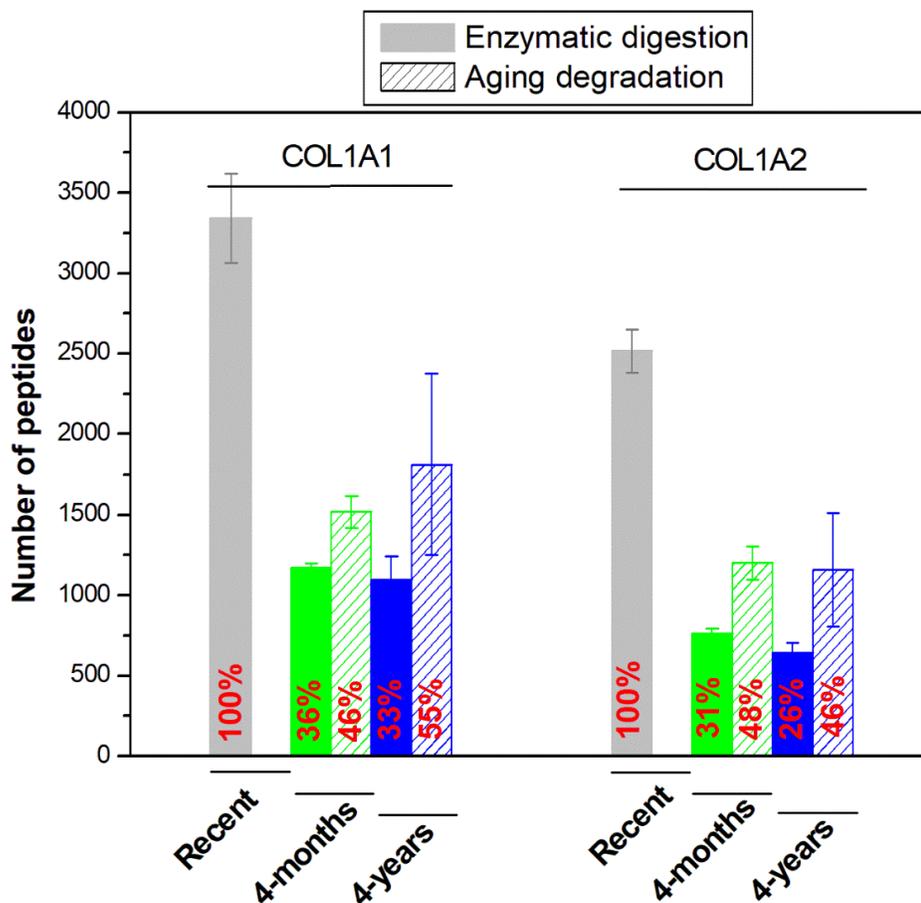


Figure 7: Number of arising peptides from enzymatic digestion and aging degradation in recent, 4-month, 4-year bovine artificial aging collagens.

3.3.3. Post-translational modifications during the aging process

Hydrophilic (*e.g.* oxidation, dioxidation, deamidation, phospho, sulfo) and hydrophobic (*e.g.* acetyl, carbamyl, carboxy, carboxymethyl, carboxyethyl, formyl, methyl) post-translational modifications were studied on the bovine and rat collagen sequences (COL1A1 and COL1A2). Differences according to the organism, age, and artificial/natural aging were observed. The number of modifications and the corresponding percentage according to the total number of possible modified sites are summarized in Table 2 and S9-11 for each sample at different ages.

Based on previous results (see section 3.3.2), about a fifth of bovine sequence information was lost during aging. The post-translational modifications studies show that the percentage of hydrophilic modifications (deamidation, phospho and sulfo) in COL1A1 decreases considerably. Exception for oxidation modifications on lysines, methionines, and prolines (exact positions in Figures S10 – S12, Tables S9, and S10), which increase in 4-months samples, before decreasing in older samples in correlation with the degradation and the loss of sequence information. On the

other hand, the polypeptide chain alpha 2 seems more subject to hydrophobic (acetyl, carbamyl, carboxymethyl, formyl, and methyl) modifications than the COL1A1, during the aging process. Exception for oxidation and phospho modifications which increase (Table 2). In sum, the decrease of hydrophilic groups and the increase of hydrophobic groups during aging may be a hypothesis to explain the evolution of the insolubility of collagen along life, to add to the protein degradations. The comparison between artificial (bovine) and natural (rat) aging on post-translational modifications on collagen sequences is summarized in Table 3. The percentages of modifications were slightly higher under artificial aging.

Table 2: Total number and percentage of post-translational modifications on bovine collagen COL1A1 and COL1A2 sequences at different artificial aging.

COL1A1							
Modifications	Amino acids	Recent (n=3)	% recovery	4- months (n=2)	% recovery	4- years (n=3)	% recovery
Oxidation	KPM	79	22,8%	98	28,2%	65	18,7%
Dioxidation	M	0	0%	0	0%	0	0%
Deamidation	NQR	5	3,4%	4	2,7%	3	2,0%
Phospho	CDKHRSTY	27	8,0%	20	5,9%	9	2,7%
Acetyl	K	0	0%	1	1,8%	0	0%
Carbamyl	K	0	0%	0	0%	0	0%
Carboxy	K	0	0%	0	0%	0	0%
Carboxymethyl	K	0	0%	0	0%	0	0%
Carboxyethyl	K	0	0%	0	0%	0	0%
Formyl	KST	9	5,3%	12	7,1%	9	5,3%
Methyl	DE	4	3,0%	3	2,2%	2	1,5%
Sulfo	STY	4	3,2%	4	3,2%	1	0,8%
COL1A2							
Modifications	Amino acids	Recent (n=3)	% recovery	4- months (n=2)	% recovery	4- years (n=3)	% recovery
Oxidation	KPM	40	13,8%	56	19,4%	40	13,8%
Dioxidation	M	0	0%	0	0%	0	0%
Deamidation	NQR	4	2,6%	3	1,9%	4	2,6%
Phospho	CDKHRSTY	9	2,9%	16	5,2%	17	5,6%
Acetyl	K	0	0%	0	0%	1	2,0%
Carbamyl	K	0	0%	1	2,0%	0	0%
Carboxy	K	0	0%	0	0%	0	0%
Carboxymethyl	K	0	0%	0	0%	1	2,0%
Carboxyethyl	K	0	0%	0	0%	0	0%
Formyl	KST	4	2,5%	6	3,8%	7	4,5%
Methyl	DE	2	1,9%	3	2,9%	3	2,9%
Sulfo	STY	1	0,8%	1	0,8%	0	0%

Table 3: Number and percentage of post-translational modifications on bovine and rat collagens COL1A1 and COL1A2 sequences at 4-months after pepsin and trypsin enzymatic treatment.

COL1A1					
Modifications	Amino acids	4-months bovine (n=2)	% recovery	4-months rat (n=2)	% recovery
Oxidation	KPM	89	25,5%	57	16,4%
Dioxidation	M	0	0%	0	0%
Deamidation	NQR	3	2%	2	1,4%
Phospho	CDKHRSTY	18	5,6%	4	1,2%
Acetyl	K	1	1,8%	0	0%
Carbamyl	K	0	0%	0	0%
Carboxy	K	0	0%	0	0%
Carboxymethyl	K	0	0%	0	0%
Carboxyethyl	K	0	0%	0	0%
Formyl	KST	8	5,0%	3	1,8%
Methyl	DE	3	2,1%	1	0,7%
Sulfo	STY	2	1,7%	0	0%
COL1A2					
Modifications	Amino acids	4-months bovine (n=2)	% recovery	4-months rat (n=2)	% recovery
Oxidation	KPM	39	13,2%	31	10,7%
Dioxidation	M	0	0%	0	0%
Deamidation	NQR	2	1,3%	2	1,3%
Phospho	CDKHRSTY	12	4,0%	5	1,6%
Acetyl	K	0	0%	0	0%
Carbamyl	K	0	0%	0	0%
Carboxy	K	0	0%	0	0%
Carboxymethyl	K	0	0%	0	0%
Carboxyethyl	K	0	0%	0	0%
Formyl	KST	4	2,7%	4	2,5%
Methyl	DE	3	2,8%	1	1,0%
Sulfo	STY	1	0,9%	1	0,8%

3.3.4. Comparison of exact positions of D-amino acids and posttranslational modifications

Both amino acid racemization and post-translational modifications can appear, simultaneously or not, during aging. Figures 8 and S14-18 regroup the exact position of D-amino acids totally racemized and the exact position of post-translational modifications.

```

1  MFSFVDLRLL  LLLAATALLT  HGQEEGQEEG  QEEDI PPVTC  VQNGRLRYHDR  50
51  DVWKPVPCQI  CVC DNGNVLC  DDVICDELKD  CPNAKVPTDE  CCFVCP EGQE  100
101 SPTDQETTGV  EGP KGD TGPR  GPRGPAG PPG  RDGIPGQ PGL  PGF PPGPP  150
151  GPPLGGNFA  PQLSYGYDEK  STGISVPGFM  GPSGPRGLPG  PPGAPGPQGF  200
201  QGPPGEFGEF  GASGPMGFRG  PGPPGKNGD  DGEAGKPGRP  GERGPPGPQ  250
251  ARGLPGTAGL  PGMKGHRGFS  GLDGAKGDAG  PAGPKGEPGS  PGENGAPGQM  300
301  GPRGLPGERG  RPGAPGPAGA  RGNDGATGAA  GPPGPTGPAG  PPGFPGAVGA  350
351  KGEGGPQGPR  GSEGPQVRG  EPGPPGAGA  AGPAGNPGAD  GQFGAKGANG  400
401  APGIAGAPGF  PGARGPSGPQ  GPSGPPGKG  NSGEPGAPGS  KGDTGAKGEP  450
451  GPTGIQGPPG  PAGEEGKRGA  RGEPGALP  GPPGERGGPG  SRGFPADGV  500
501  AGPKGPAGER  GAPGPAGPKG  SPGEAGRPGE  AGLPGAKGLT  GSPGSPGPDG  550
551  KTGPPGAGQ  DGRPGPP  GARGQAGVM  FPGPKGAGE  PGKAGERGVP  600
601  GPPGAVPAG  KDGEAGAQGP  PGPAGPAGR  GEQGPAGSPG  FQGLPGAP  650
651  PGEAGKPGEQ  GVPGDLGAPG  PSGARGERGF  PGERGVQGPP  GPAGPRGANG  700
701  APGNDGAKD  AGAPGAPGSQ  GAPGLQGMPG  ERGAAGLGP  KGDRGDAGPK  750
751  GADGAPKDG  VRGLTGPIG  PGPAGAPGDK  GEAGPSGPAG  PTGARGAPDG  800
801  RGEPPGAPA  GFAGPGADG  QPGAKGEPD  AGAKGDAGPP  GPAGPAGPPG  850
851  PIGNVGAPG  KGARGSAGPP  GATGFPGAAG  RVGPPPSGN  AGPPPGAPA  900
901  GKEGSKPRG  ETGPARPE  VGPPGPGA  GEKGAPGADG  PAGAPGTPG  950
951  QGIAGQRGVV  GLPQRGERG  FPGLPGSGE  PGKQGPSGAS  GERGPPGMG  1000
1001  PPLAGPPE  SGREGAPGA  GSPGRDGSPG  AKGDRGETGP  AGFPAPGAP  1050
1051  GAPGPVPAG  KSGDRGETGP  AGPAGPIGPV  GARGPAGPQG  PRGDKGETGE  1100
1101  QGDRGIKHR  GFSGLQGPPG  PGSPGEQGP  SGASGPAGPR  GPPGSAGSPG  1150
1151  KDGLNGLPG  IGPPPRGRT  GDAGPAGPPG  PPGPPPGP  PSGGYDLSFL  1200
1201  PQPQEKAHD  GGRYRADDA  NVVRDRDLEV  DTTLKSLSQ  IENIRSPEGS  1250
1251  RKNPARTCRD  LKMCHSDWKS  GEYWIDPNQG  CNLDAIKVFC  NMETGETCVY  1300
1301  PTQSVAQKN  WYISKNPKEK  RHVWYGESMT  GGFQFEYGGQ  GSDPADVAIQ  1350
1351  LTFLRLMSTE  ASQNITYHCK  NSVAYMDQQT  GNLKKALLLQ  GSNEIEIRAE  1400
1401  GNSRFTYSVT  YDGCTSHTGA  WGKTVIEYKT  TKTSRLPIID  VAPLDVGAPD  1450
1451  QEFGFDVGPA  CFL  1463

```

Figure 8: 4-years bovine collagen COL1A1 sequence modifications: post-translational modifications (red), amino acids in their D-form totally racemized (blue), and both modifications on the same amino acid (green).

Conclusion

Artificial aging was successfully applied to standard collagen to mimic the natural aging in biological samples. Several collagen samples from different organisms (bovine and rat), were studied at different ages (recent, 4-months, 4-years). DCI/D₂O hydrolysis was privileged to HCl/H₂O to limit the natural racemization of amino acids during the protein hydrolysis. This new

chiral analysis method allowed us to determine the percentage of D-amino acids in collagen according to age. Results presented show a % D-amino acids – age correlation. Peptide analysis shows that the amount of peptides from aging degradation is higher than arising peptides from enzymatic digestions. The collagen protein sequence is increasingly degraded over time. The post-translational modifications study showed a decrease of hydrophilic groups and an increase of hydrophobic groups during aging. These sequence modifications may be a hypothesis to explain the evolution of the insolubility of collagen throughout life. The percentage of post-translational modifications was not significantly different between artificial and natural aging in both organisms. The combination of these results allowed us to determine the exact positions of D-amino acids and PTMs.

CRedit authorship contribution statement

Marine Morvan: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – Review & Editing, Visualisation. **Ivan Mikšik:** Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Czech Science Foundation, grant number 20-03899S.

References

- [1] R. L. Levine, E. R. Stadtman, Oxidative modification of proteins during aging, *Experimental Gerontology* 36 (2001) 1495-1502.
- [2] G. Krishnamoorthy, R. Selvakumar, T. P. Sastry, A. B. Mandal, M. Doble, Effect of d-amino acids on collagen fibrillar assembly and stability: Experimental and modelling studies, *Biochemical Engineering Journal* 75 (2013) 92–100.
- [3] J. W. Silzel, G. Ben-Nissan, J. Tang, M. Sharon, R. R. Julian, Influence of Asp Isomerization on Trypsin and Trypsin-like Proteolysis, *Anal. Chem.* 94 (2022) 15288–15296.
- [4] A. L. Santos, A. B. Lindner, Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease, *Oxidative Medicine and Cellular Longevity* 2017 (2017) 1–19.

- [5] I. Mikšík, P. Sedláková, S. Pataridis, F. Bortolotti, R. Gottardo, Proteins and their modifications in a medieval mummy: Proteomic Analysis of Mummy, *Protein Science* 25 (2016) 2037–2044.
- [6] R. J. W. Truscott, J. Mizdrak, M. G. Friedrich, M. Y. Hooi, B. Lyons, J. F. Jamie, M. J. Davies, P. A. Wilmarth, L. L. David, Is protein methylation in the human lens a result of non-enzymatic methylation by S-adenosylmethionine?, *Experimental Eye Research* 99 (2012) 48–54.
- [7] S. Sh. Atavliyeva, P. V. Tarlykov, Paleoproteomics studies of ancient caprinae: a review 89 (2021) 4-14.
- [8] N. Fujii, T. Kawaguchi, H. Sasaki, N. Fujii, Simultaneous Stereoinversion and Isomerization at the Asp-4 Residue in β B2-Crystallin from the Aged Human Eye Lenses, *Biochemistry* 50 (2011) 8628–8635.
- [9] M. Y. S. Hooi, R. J. W. Truscott, Racemisation and human cataract. d-Ser, d-Asp/Asn and d-Thr are higher in the lifelong proteins of cataract lenses than in age-matched normal lenses, *AGE* 33 (2011) 131–141.
- [10] M. Y. S. Hooi, M. J. Raftery, R. J. W. Truscott, Accelerated aging of Asp 58 in α A crystallin and human cataract formation, *Experimental Eye Research* 106 (2013) 34–39.
- [11] B. Lyons, A. H. Kwan, J. Jamie, R. J. W. Truscott, Age-dependent modification of proteins: N-terminal racemization, *FEBS J* 280 (2013) 1980–1990.
- [12] N. Fujii, T. Takata, N. Fujii, K. Aki, H. Sakaue, D-Amino acids in protein: The mirror of life as a molecular index of aging, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866 (2018) 840–847.
- [13] S. Ha, T. Kinouchi, N. Fujii, Age-related isomerization of Asp in human immunoglobulin G kappa chain, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1868 (2020) 140410.
- [14] M. Morvan and I. Mikšík, Recent Advances in Chiral Analysis of Proteins and Peptides, *Separations* 8 (2021) 112.
- [15] N. Fujii, T. Takata, I. Kim, K. Morishima, R. Inoue, K. Magami, T. Matsubara, M. Sugiyama, T. Koide, Asp isomerization increases aggregation of α -crystallin and decreases its chaperone activity in human lens of various ages, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1868 (2020) 140446.
- [16] V. V. Dyakin, T. M. Wisniewski, A. Lajtha, Racemization in Post-Translational Modifications Relevance to Protein Aging, Aggregation and Neurodegeneration: Tip of the Iceberg, *Symmetry* 13 (2021) 455.
- [17] D. Vicente-Zurdo, S. Rodríguez-Blázquez, E. Gómez-Mejía, N. Rosales-Conrado, M. E. León-González, Y. Madrid, Neuroprotective activity of selenium nanoparticles against the effect of amino acid enantiomers in Alzheimer's disease, *Anal. Bioanal. Chem.* 414 (2022) 7573–7584.
- [18] A. A. Ageeva, A. B. Doktorov, N. E. Polyakov, T. V. Leshina, Chiral Linked Systems as a Model for Understanding D-Amino Acids Influence on the Structure and Properties of Amyloid Peptides, *IJMS* 23 (2022) 3060.
- [19] P. M. Helfman, J. L. Bada, Aspartic acid racemization in tooth enamel from living humans, *Proc. Natl. Acad. Sci. U.S.A.* 72 (1975) 2891–2894.
- [20] J. T. Powell, N. Vine, M. Crossman, On the accumulation of d-aspartate in elastin and other proteins of the ageing aorta, *Atherosclerosis* 97 (1992) 201–208.

- [21] S. Ritz-Timme, I. Laumeier, M. J. Collins, Aspartic acid racemization: evidence for marked longevity of elastin in human skin, *Br. J. Dermatol.* 149 (2003) 951–959.
- [22] J. M. Lee, L. Petrucelli, G. Fisher, S. Ramdath, J. Castillo, M. M. Di Fiore, A. D'Aniello, Evidence for D-Aspartyl- β -Amyloid Secretase Activity in Human Brain, *J. Neuropathol. Exp. Neurol.* 61 (2002) 125–131.
- [23] S. Ritz, A. Turzynski, H. W. Schütz, A. Hollmann, G. Rochholz, Identification of osteocalcin as a permanent aging constituent of the bone matrix: basis for an accurate age at death determination, *Forensic Science International* 77 (1996) 13–26.
- [24] G. Yasunaga, S. Inoue, T. Bando, T. Hakamada, Y. Fujise, Aspartic acid enantiomer quantification using ultraperformance liquid chromatography–tandem mass spectroscopy combined with deuterium-chloride hydrolysis to improve age estimation in Antarctic minke whale *Balaenoptera bonaerensis*, *Marine Mammal Science* (2022) 1-19.
- [25] N. S. Mahlke, S. Renhart, D. Talaa, A. Reckert, S. Ritz-Timme, Molecular clocks in ancient proteins: Do they reflect the age at death even after millennia?, *Int. J. Legal. Med.* 135 (2021) 1225–1233.
- [26] G. T. Matteussi, V. Jacometti, A. Franco, R. H. A. da Silva, Age estimation in humans through the analysis of aspartic acid racemization from teeth: A scoping review of methods, outcomes, and open research questions, *Forensic Science International* 331 (2022) 111154.
- [27] H. Frank, W. Woiwode, G. Nicholson, E. Bayer, Determination of the rate of acidic catalyzed racemization of protein amino acids, *Liebigs Ann. Chem.* 1981(1981) 354-365.
- [28] J. Csapá, Z. Csapó-Kiss, L. Wágner, T. Tálos, T. G. Martin, S. Folestad, A. Tivesten, S. Némethy, Hydrolysis of proteins performed at high temperatures and for short times with reduced racemization, in order to determine the enantiomers of d- and l-amino acids, *Analytica Chimica Acta* 339 (1997) 99–107.
- [29] T. Miyamoto, H. Homma, Detection and quantification of d -amino acid residues in peptides and proteins using acid hydrolysis, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866 (2018) 775–782.
- [30] M. Danielsen, C. Nebel, T. K. Dalsgaard, Simultaneous Determination of L- and D-Amino Acids in Proteins: A Sensitive Method Using Hydrolysis in Deuterated Acid and Liquid Chromatography–Tandem Mass Spectrometry Analysis, *Foods* 9 (2020) 309.
- [31] R. Pérez-Míguez, B. Bruyneel, M. Castro-Puyana, M. L. Marina, G. W. Somsen, E. Domínguez-Vega, Chiral Discrimination of DL-Amino Acids by Trapped Ion Mobility Spectrometry after Derivatization with (+)-1-(9-Fluorenyl)ethyl Chloroformate, *Anal. Chem.* 91 (2019) 3277–3285.
- [32] T. Furuhashi, W. Weckwerth, Isomer analysis by mass spectrometry in clinical science, *TrAC Trends in Analytical Chemistry* 153 (2023) 116907.
- [33] S. Ricard-Blum, The Collagen Family, *Cold Spring Harbor Perspectives in Biology* 3 (2011) a004978–a004978.
- [34] S. Köster, H. M. Evans, J. Y. Wong, T. Pfohl, An In Situ Study of Collagen Self-Assembly Processes, *Biomacromolecules* 9 (2008) 199–207.
- [35] N. K. Shah, B. Brodsky, A. Kirkpatrick, J. A. M. Ramshaw, Structural consequences of D-amino acids in collagen triple-helical peptides, *Biopolymers* 49 (1999) 297–302.
- [36] V. Punitha, S. Sundar Raman, R. Parthasarathi, V. Subramanian, J. Raghava Rao, Balachandran Unni Nair, and T. Ramasami, Molecular Dynamics Investigations on the Effect of D Amino Acid Substitution in a Triple-Helix Structure and the Stability of Collagen, *J. Phys. Chem. B* 113 (2009) 8983–8992.

[37] A. Vrána, L. Kazdová, The hereditary hypertriglyceridemic nonobese rat: an experimental model of human hypertriglyceridemia, *Transplant. Proc.* 22 (1990) 2579.