Identification of collagen 1α3 in teleost fish species and typical collision induced internal fragmentations

Anne J. Kleinnijenhuis a,*, Frédérique L. van Holthoon b, Bastiaan van der Steen b

a Triskelion, Reactorweg 47A, 3542 AD Utrecht, The Netherlands
b Rousselot Expertise Center, Meulesteedkaai 81, B-9000 Ghent, Belgium

A R T I C L E   I N F O
Keywords:
Collagen 1α3
Teleost
Fish
LC-MS
Gelatin
Internal fragmentation

A B S T R A C T
In contrast to collagens 1α1 and 1α2, the more obscure collagen 1α3 is sparsely mentioned in literature. In skin collagen type 1 of teleosts (bony fish), however, the chain occurs in a heterotrimer together with collagens 1α1 and 1α2, which makes it one of the most abundant proteins in teleosts. As teleost fish species and gelatin (hydrolysate) prepared from their skin are a major source for food products and nutraceuticals, the goal of the study was to selectively identify collagen 1α3 in several fish species. Fish skin extracts and fish skin gelatins were analyzed using LC-MS. Depending on the amount of available genetic information different approaches were used to identify collagen 1α3. Additionally, collagen-specific collision induced internal fragmentations are discussed, which are important to consider during data analysis. Ultimately the presence of collagen 1α3 could be confirmed using LC-MS in multiple fish species.

Introduction

Collagens are basic structural proteins (Karsdal, Leeming, Henriksen, & Bay-Jensen, 2016) in both vertebrates and invertebrates (Shahidi, 2007). They provide mechanical stability, strength and toughness to several tissues such as tendon, skin and bone (Fratzl, 2008). The protein family has a long history, originating from before the Cambrian period. Collagen is closely involved with the appearance and evolution of metazoa (Exposito, Cluzel, Garrone, & Lethias, 2002): all animals that undergo development from an embryonic stage with three tissue layers (ectoderm, mesoderm and endoderm) (Technau, & Scholz, 2003). Amino acid sequence comparisons of various collagens indicate that the main types of collagen evolved about 800–900 million years ago (Runnegar, 1985). Chaoanoflagellate and diploblast genomic data have suggested that the formation of an ancestral α chain occurred before the metazoan radiation (King et al., 2008; Zhang et al., 2007). Phylogenetic studies point to an early emergence of the three fibrillar collagen clades A, B and C before the metazoan radiation (Exposito, Valcourt, Cluzel, & Lethias, 2010). While clade B comprises collagen types 5α1, 5α3, 11α1 and 11α2 and clade C collagen types 24 and 27, the clade A collagens include types 1 to 3 and 5α2 (Exposito et al., 2008; Boot-Handford, Tuckwell, Plumb, Rock, & Poulson, 2003; Sicot et al., 1997). The classification and major characteristics of fibrillar collagens are summarized in Table 1. Clade A collagen type 1 is by far the most abundant protein in vertebrates (Makeeva, & Leikin, 2014) and forms type 1 triple helices (Shoulders, & Raines, 2009). Being formed from procollagen, the triple helices usually consist of heterotrimers of one collagen 1α1 and two collagen 1α1 chains (Bellamy, & Bornstein, 1971); therefore, the latter chains are relatively well known. The more obscure collagen 1α3 chain, which is the main subject of this paper, occurs only in teleosts (bony fish). Skin type 1 collagens of several teleosts contain 1α3 in a heterotrimer, together with 1α1 and 1α2, as shown for Alaska pollack (Kimura, & Ohno, 1987; Piez, 1965) and common mackerel (Kimura, 1985). From a phylogenetic point of view, it seems likely that the collagen 1α3 gene emerged around the time of the adaptive radiation of bony fish (Kimura, Ohno, Miyauchi, & Uchida, 1987). This is the reason why collagen 1α3 does not occur in other fish species, amphibians, reptiles, birds or mammals. Results obtained in zebrafish support the hypothesis that the 1α3 chain arose from a duplication of the 1α1 gene (Morvan-Dubois, Le Guelliec, Garrone, Zylberberg, & Bonnoud, 2003) and the original and duplicate genes have diverged ever since. To distinguish and assign collagen 1α1 and 1α3 protein sequences in a fish species, ideally complete genetic information is available of these proteins. When this information has not or only partly been uncovered for a species, it is difficult to link elucidated sequences to either collagen 1α1 or 1α3, solely based on LC-MS/MS (Liquid Chromatography - tandem LC-MS).
mass spectrometry) data. In these cases, it can be helpful to search the LC-MS/MS data against species closely related to the analyzed species. Another challenging aspect is to find reliable collagen 1α3 database information at all, as often a different nomenclature is used, namely collagen 1α1b. In the NCBI nucleotide database few sequences can be found which have collagen 1α3 in the accession title, but several more which have collagen 1α1b in the title. During our search for collagen 1α3 sequences it appeared that collagen 1α3 is also described as “collagen 1α1-like” in some occasions. Finally, another consideration is the polyplody that several fish species exhibit (Leggatt, & Iwama, 2003) which may complicate the naming of proteins, although it would not hinder the identification of collagen 1α3 chains in fish species from a data analytical point of view.

Collagen 1α3/1α1b has been detected in proteomic and other studies (Keen et al., 2018; Carlson, Smalley, & Van Beneden, 2013; Koth et al., 2020), but it is rarely mentioned in literature. Considering its very high abundance in teleosts, having the same stoichiometry as the 1α1 and 1α2 chains in skin, it is an interesting protein for further investigation. Collagen type 1 is a basic constituent of food and an ingredient to produce several food and nutraceutical products, such as gelatin and collagen hydrolysate. Commercial gelatins are mainly produced from bovine and porcine skins and bones, and consequently these gelatins are mainly composed of partly hydrolyzed type 1 collagen (Stevens, 2010; Vergauwen et al., 2016; Yasui, Ono, Konomi, & Nagai, 1984; Kleinnijenhuis, Van Holthoon, & Herregods, 2018). Another source to produce gelatin is the skin of several fish species (Vergauwen et al., 2016). Gelatin is typically further chemically and/or enzymatically processed to produce collagen hydrolysate nutraceutical products. When it is considered that teleost fish species are a major food source and that several food and nutraceutical products are prepared from their skins, it can be stated that collagen 1α3 is an important protein worth further investigation.

The source protein composition of many mammalian collagen (hydrolysate) products is quite well known as they contain collagen 1α1 and 1α2 in a 2:1 molar ratio. In addition, it is often known which other collagen types might be present in raw material tissues, such as collagen 2α1 (abundant in cartilage) or collagen 3α1 (abundant in skin). To shed more light on the composition of fish gelatin (hydrolysates) we decided to selectively investigate the presence of collagen 1α3 in skin extracts and skin gelatin samples of several fish species, using ultra high performance LC-MS/MS (UHPLC-MS/MS) with an Orbitrap analyzer. It is essential to use a high-resolution mass analyzer during collagen MS/MS sequence analysis, to discriminate between hydroxyproline and leucine/isoleucine residues, which differ only 0.036 Da in mass. Fish species or families can often be readily assigned using elucidated collagen 1α1 and 1α2 sequences, but the more challenging goal of this study was to identify collagen 1α3. The examined fish species were categorized based on the amount of available genetic information, and the approach for finding collagen 1α3 was adapted to this aspect. After having obtained a global overview of the similarities and relations between clade A fibrillar collagens, the LC-MS investigation of fish skin collagens is discussed, starting with species for which complete genetic collagen 1α3 information was available to species with hardly any information. Additionally, collision induced dissociation (CID) of tryptic collagen peptides is discussed, featuring collagen-specific typical internal fragmentations. Ultimately the presence of collagen 1α3 could be confirmed using LC-MS in multiple fish species.

### Materials and methods

**Gelatin samples and collagen extraction**

Type A (acid pre-treated) fish skin gelatin samples ( pangasius, tilapia, cod and salmon) were kindly provided by Rousselot. Other fish (barramundi, sea bream, trout, hake, cod, saithe/pollack, mackerel, rose fish, haddock, salmon, sea bass and sardine) were purchased at local supermarkets. Collagen from these fish was extracted from the skins based on the procedure described by Gudmundsson and Hafsteinsson (Gudmundsson, & Hafsteinsson, 1997) and Koli et al. (Koli, Basu, Gudipati, Chouksey, & Nayak, 2013). In short, skins (approximately 1 g) were cleaned and rinsed with water to remove excess material and treated for 40 min respectively with 0.2% (w/v) sodium hydroxide, 0.2% (w/v) sulfuric acid and 1.0% (w/v) citric acid solutions. After each treatment, the skins were washed under running tap water until the pH was between 6.5 and 7.5. The soaking and washing treatments were repeated three times. Finally, collagen was extracted from skin in 7 ml milliQ water at 50 °C for 18 h. After solubilization, reduction, alkylation and tryptic digestion the samples were analyzed using UHPLC-MS/MS.

**LC-MS analysis**

Samples were analyzed using a combination of a UHPLC (Ultimate 3000, Dionex) and a Q-Exactive mass spectrometer (ThermoElectron). An Acuity HSS T3 column (2.1 × 100 mm, 1.8 μm, Waters, Milford, PA, USA) was used at a temperature of 40 °C. Elution was achieved using a binary gradient from 2% to 30% B at a flow rate of 0.5 ml minute⁻¹ with solvents A and B, both containing 0.1% formic acid in respectively milliQ water and acetonitrile, followed by a column wash and equilibration as part of the gradient, to prepare for the next run. The autosampler temperature was 40 °C and the injection volume was 10 μl. The total run time was 18 min. All peptides were analyzed using electrospray ionization in positive mode (HESI source) using a full-scan data-dependent mode. The mass range was set to m/z 200–2000 at a resolution of 35000. The top 5 ions were submitted to data-dependent scans at a normalized collision energy of 15, 25 and 35. The spray voltage was 3.0 kV. Other settings were: sheath gas flow (60 AU), auxiliary gas flow (20 AU), capillary temperature (320 °C), heater temperature (350 °C), S-Lens RF Level (50 V), AGC target (1e6), and maximum IT (150 ms). XCalibur software version 3 (ThermoScientific) was used for data acquisition.

**Data analysis**

Database analysis of the raw files was performed in Proteome Discoverer 1.4 (Thermo) and included manual inspection/confirmation. Sequences of 48 collagen 1α3 or 1α1b mRNA or cDNA entries were obtained from the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore/advanced), accessed in April–June 2021 (see Supplementary Table 1). The sequences were translated to protein to compose a fasta file. Collagen 1α3/1α1b sequences were only included in the fasta file if the GXY domain was 1014 codons long, to promote the inclusion of sequences with high quality (Kleinnijenhuis, 2019). An overview of other collagen data sources is provided in Supplementary Table 2. From the Clupea harengus sequence XP_031416510.1 one unidentified amino acid was deleted as it disrupted the GXY pattern. The relatedness of sequences was assessed using Clustal Omega analysis (Sievers et al., 2011).

Database searching parameters included full tryptic digestion and
Results and discussion

To visualize the similarities and relations between fibrillar collagens of clade A, the number of nucleotide differences between the GXY domain cDNA sequences was calculated for collagens 1–3 of cow, pig, chicken, coelacanth, and tilapia. The numbers of mutual nucleotide differences were set out in a matrix and collagen 1\alpha1/1b sequences of several fish species were added, see Fig. 1. The accessions of the compared species are mentioned in Supplementary Table 1 and Supplementary Table 2. Collagen 5\alpha2 was excluded from this comparison due to its low similarity to collagens 1–3 (results not shown). Not every collagen type is represented for the selected animal species in Fig. 1, because specific collagen types do not occur in an animal species or due to absent information. It should be noted that the matrix, showing similarities on the nucleotide level, does not linearly translate to similarity on the amino acid level or to evolutionary divergence time and rate (Kleinnijenhuis, 2019). Analysis of Fig. 1 resulted in the following findings:

a) The relations between fish 1\alpha3 and 1\alpha1b sequences versus the relations with the collagen 1\alpha1 sequences indicate that the chain names 1\alpha3 and 1\alpha1b are indeed synonyms. The same conclusion was drawn from a similar extended comparison of all the accessions mentioned in Supplementary Table 1 (results not shown) and a Clustal Omega analysis of all the sequences (translated to protein) from Supplementary Table 1 and part of the sequences in Supplementary Table 2, see Supplementary Fig. 1. In the remainder of the document the chain of interest will be denoted as collagen 1\alpha3.

Additionally, the Clustal Omega analysis revealed that the selected collagen 1\alpha1 sequences formed an individual cluster within the set of investigated sequences.

b) The similarities between the collagen 1\alpha3 and 1\alpha1 sequences indicate that collagen 1\alpha3 did not emerge long before teleost radiation, confirming earlier findings (Kimura, Ohno, Miyauchi, & Uchida, 1987). Therefore, it is difficult to discriminate between collagen 1\alpha1 and 1\alpha3 in fish species when there is limited genetic or protein sequence information available.

c) On the cDNA level, collagen 1\alpha3 and 1\alpha1 are most similar to collagen 2\alpha1, followed by 1\alpha2 and 3\alpha1.

d) The added cod (Gadus morhua) collagen 1\alpha1-like sequence XM_033397920.1, which will be discussed in more detail below, is more similar to the collagen 1\alpha3 sequences than to the collagen 1\alpha1 sequences. It is probable that this cod collagen 1\alpha1-like sequence thus represents collagen 1\alpha3, which makes it reasonable to explore the feasibility of using collagen 1\alpha1-like sequences in the search for collagen 1\alpha3.

Due to finding b) it was decided to sort the analyzed fish species in categories I to IV. The categorization was based on the available collagen 1\alpha3 sequence information: I) sequence information available of the species (tilapia), II) sequence information available of family member (trout, sea bream, rose fish, salmon, sea bass, pangasius), III) sequence information available of order member (barramundi, mackerel) and IV) no sequence information available of order member (hake, cod, saithe, haddock, sardine). Per category different approaches were used to identify collagen 1\alpha3. To assign collagen 1\alpha3, software-based data analysis was applied, followed by manual confirmation. Due to the collagen-specific internal fragment ion series that are typically observed, software-based data analysis results must be carefully assessed. Additionally, wrong assignment of the location of hydroxyproline (Hyp or p) residues is often observed. The potential of incomplete sequence information in MS/MS and wrongly assigned modification sites in relation to permutations and isomeric combinations, together with erroneous assignment of internal fragment ions as...
fragment ions containing a terminus, can be problematic. These issues especially apply to collagens as their amino acid sequences contain many slightly different repetitions (Kleinnijenhuis et al., 2020). Besides extensive manual inspection of MS/MS data, it is an option to improve proteomic algorithms or models (Yang et al., 2021), but the specific extensive manual inspection of MS/MS data, it is an option to improve many slightly different repetitions (Kleinnijenhuis et al., 2020). Besides especially apply to collagens as their amino acid sequences contain

The only category I fish species (sequence information available of the species) was tilapia (Oreochromis niloticus). The aligned GXY domain amino acid sequences of tilapia collagen 1α1, 1α2, 1α3 and 2α1 are shown in Fig. 2. The highlighted trypsin cleavage sites exemplify the similarity between the chains. It was decided to focus the analysis on a part of the chains, which has highly conserved (preceding) trypsin cleavage sites, also considering their location in other animal species and collagen types mentioned in Supplementary Table 1 and Supplementary Table 2. Between the (preceding) trypsin cleavage sites, however, this part of the chain does exhibit amino acid changes between collagen types of the same species, which is important to discriminate collagen 1α3 and 1α1. The part of the sequence is highlighted as bold in Fig. 2 and represents residues 220–237 of the GXY domain. A combined ion chromatogram of the analogous peptides from tilapia collagen 1α1, 1α2 and 1α3 is presented in Fig. 3. Although the chromatogram shows qualitative results, the analogous peptides are expected to have a relatively similar response factor as their sequences are similar, and therefore the MS results indicate the expected 1:1:1 stoichiometry of the three chains in tilapia skin. The sequences of the tilapia peptides were confirmed by MS/MS, see Fig. 4, which also illustrates the collagen-specific internal fragmentations. Like proline (Breci, Tabb, Yates, & Wysocki, 2003), hydroxyproline exhibits intense collision induced backbone fragmentations on the N-terminal side of the residue. A majority of the collagen trypsin cleavage sites is located C-terminally of the third GXY domain sequence. Collagen 1α3 peptide including residues 220–237) are shown in Fig. 4. The CID spectrum of the analogous tilapia collagen 1α2 peptide, which is not a Hyp-3 peptide, does not show the typical Hyp-3 related internal fragment ions, but relatively more b type ions, which contain the N-terminus. The spectrum does contain internal fragment ions starting with pG*, but they originate from the Hyp residue at the sixth position. Internal fragment ion series for Hyp-6 or Hyp-9 peptides are observed regularly, but with lower intensity than for Hyp-3 peptides, which indicates that especially the loss of the two N-terminal amino acids is favorable when Hyp is in the third position. Overall, it was concluded that tryptic collagen peptide CID fragmentation is strongly influenced by the presence of Hyp, especially when the modified amino acid is in the third position. Collagen 1α3 was clearly identified in tilapia. The elucidated tilapia peptide sequences did not provide hits with other tilapia proteins when searched against taxid (taxonomic identifier) 8128 (Oreochromis niloticus) using protein blast, which further supports the assignments. All the collagen 1α3 identifications are summarized in Table 2; it should be specifically mentioned that often the commercial names of fish, sold as food or used as a basis for food products, do not directly relate to single species names as different species are sold under a common name (e.g. salmon), or could even represent mixtures of species (e.g. sardine). The peptide sequences in Table 2 were confirmed by assessing (low and high intensity) b and y ions as well as Hyp-3 peptide internal fragments including pG*.

For several other analyzed fish species, it was not possible to find the tryptic collagen 1α3 containing residues 220–237 of the GXY domain. In these cases it was decided to resort to a different peptide, under the condition that the analogous tilapia collagen 1α3 and 1α1 peptides had the same (preceding) trypsin cleavage locations and that there were at least two mutual amino acid differences between the

---

**Table 1.** Collagen type and Aligned collagen helix sequence for tilapia collagen 1α1, 1α2, 1α3 and 2α1. Tryptsin cleavage sites are in blue (K) and green (R). Proline residues C-terminal of K and R, resulting in tryptic miscleavage, are in orange. The MS/MS spectra of the peptides in bold (at position 220–237) are shown in Fig. 4.

| Collagen type | Aligned collagen helix sequence | position |
|--------------|-------------------------------|----------|
| 1α1          | **GAAGIAGA**                  | 150      |
| 1α2          | **GAAGIAGA**                  | 180      |
| 1α3          | **GAAGIAGA**                  | 300      |
| 2α1          | **GAAGIAGA**                  | 450      |
| 2α2          | **GAAGIAGA**                  | 1014     |

**Fig. 2.** Aligned GXY domain sequences of tilapia (Oreochromis niloticus) collagen 1α1, 1α2, 1α3 and 2α1. Tryptsin cleavage sites are in blue (K) and green (R). Proline residues C-terminal of K and R, resulting in tryptic miscleavage, are in orange. The MS/MS spectra of the peptides in bold (at position 220–237) are shown in Fig. 4.
analogous tilapia collagen 1α3 and 1α1 peptides. It was assumed that when peptides were matched exactly to a collagen 1α3 sequence using these criteria, this indicated the presence of collagen 1α3 with high probability. Finally, elucidated peptide sequences of analyzed fish species were searched against the relevant species, family and/or order using protein blast, see Table 2. Of course, the number of hits after this type of search depends heavily on the proteome content of a database.

The category II species (sequence information available of family member) were trout, sea bream, rose fish, salmon, sea bass and pangasius. Especially for category II species the search of MS/MS data against the composed collagen 1α3 fasta file was helpful. As the fasta file contained information of family members of category II species, the search provided many hits with high confidence, which were used as a basis for confirmation of collagen 1α3 peptides. Collagen 1α3 could be identified in all the category II species and the peptide structures which were confirmed, are summarized in Table 2. Protein blast did not return relevant hits with other proteins (of the same species). Only for sea bream a collagen 2α1-like hit was obtained with Sparus aurata, see Table 2, which appears to be collagen 1α1-like as the gene names in the entries of the 1α1-like and 2α1-like hits were the same. Imprecise protein annotations are a complicating factor during data analysis. In Supplementary Fig. 2, MS/MS spectra are presented for two of the category II species, pangasius and salmon, ofryptic peptides also originating from position 220–237 of the GXY domain, confirming the presence of collagen 1α3 in the skin of these fish species. The pangasius peptide is a Hyp-3 peptide, again showing the typical internal fragment ion series, whilst the salmon peptide does not clearly show the typical fragmentation as it is not a Hyp-3 peptide. The elucidated pangasius peptide sequence did not provide hits with other proteins when searched against taxid 7999 (Pangasiidae) using protein blast. The salmon peptide gave hits with collagen 1α1b and collagen 1α1-like of several salmon and trout species when searched against taxid 8015 (Salmonidae). In addition hits were obtained with collagen 1α1 of Salvelinus alpinus (XP_023849227.1) and Oncorhynchus kisutch (XP_020316038.1), which is peculiar. The latter protein sequences were subjected to Clustal Omega analysis, together with a control Oncorhynchus mykiss collagen 1α1 sequence (NM_001124177.1) in addition to the set reported in Supplementary Fig. 1. The Salvelinus alpinus (XP_023849227.1) and Oncorhynchus kisutch (XP_020316038.1) sequences, annotated as collagen 1α1, clustered with collagen 1α3 sequences, while the control Oncorhynchus mykiss collagen 1α1 sequence (NM_001124177.1) clustered with other collagen 1α1 sequences (results not shown). This finding is another indication that collagen annotations are not always precise. Due to the obtained collagen 1α3 hits for the salmon peptide GSTGAAGISGApGFpGTR with several Salmonidae, a different target should be used to distinguish salmon from related trout species. Overall, it is sufficient for the confirmation of the presence of collagen 1α3. Again, the goal of this study was not to show how to discriminate (closely related) species, but to identify collagen 1α3 in fish species. When selectively investigating the presence of salmon and related species, it is advisable to select a less conserved target.

The category III species (sequence information available of order member) were barramundi and mackerel. For these species it was harder to identify collagen 1α3. This is illustrated in Table 3, which summarizes the highest sequence coverages per analyzed fish sample and the corresponding species in the fasta file. For the analyzed species without a family member in the fasta file there were few peptide hits with high confidence. Fortunately, the barramundi and mackerel data gave suitable hits with other, less related species and therefore collagen 1α3 could be identified and confirmed, see Table 2. From the barramundi and mackerel results it was deduced that sequence information of order members is substantially less suitable than information of family members, which is not surprising due to the potentially long divergence times between order members.

The category IV species (no sequence information available of order member) were hake, cod, saithe, haddock and sardine. Considering the relatively low number of suitable hits even for the category III species, it could be expected that the category IV species would not give any suitable hits which is illustrated in Table 3. The low sequence coverages usually consist of relatively short peptides, which are therefore not unique and occur with the same sequence across species and even across collagen types. Although initially only sequences with 1α3 and 1α1b in

![Combined extracted chromatogram of analogous peptides from tilapia collagen 1α1 (GApGAAGVAGApGFpGPR), collagen 1α2 (GAAGTpVGAGApGFpGPR) and collagen 1α3 (GApGAAIGAgApGFpGAR), present in tilapia skin gelatin.](image-url)
Fig. 4. MS/MS spectra of tryptic tilapia collagen 1 peptides from tilapia skin gelatin; a) 1α1, b) 1α2 and c) 1α3.
Table 2
Manually confirmed collagen 1α3 peptide identifications. In the third column the protein blast results are summarized.

| Animal species (category) | Elucidated peptide | Search against, description 100% hits |
|---------------------------|---------------------|----------------------------------------|
| Barramundi (III)          | GPAGAQGGVGAPGPK     | Percifomes (taxid 8111), no hits.       |
| Sea bream (II)            | EGSGQGHDApGR        | Percifomes (taxid 8111), no hits.       |
| Trout (II)                | GSTGAAGISGAPGpGTR   | Salmoniformes (taxid 8006), hit with collagen 1α1-like of several trout and salmon species. Hit with collagen 1α1b of Oncorhynchus mykiss (taxid 8022), but not with collagen 1α1 of Oncorhynchus mykiss. |
| Sardine (IV)              | GATGSpIGAPGpGFPpGR  | Clupeiformes (taxid 32446), hit with collagen 1α1-like of Clupea harengus (XP_031416510.1). |
| Hake (IV)                 | EGSTGHDApGR         | Gadiformes (taxid 8043), hit with collagen 1α1-like of Gadus morhua (XP_030195580.1 (same as translated sequence XM_03039720.1)). |
| Cod (IV)                  | GPAGAQGLGLApGPK     | Gadiformes (taxid 8043), hit with collagen 1α1-like of Gadus morhua (XP_030195580.1). |
| Saithe (IV)               | GPAGAQGPLGLApGPK    | See cod.                                |
| Mackerel (III)            | GGAGpIGATGpGpAPGR   | Percifomes (taxid 8111), hits with collagen 1α1b of Sebastes umbrosus, Epinephelus lanceolatus and Cyclopterus lumpus. |
| Rose fish (II)            | GGGpGSIGAPGpGpGSR   | Scorpifomes (taxid 8111), hit with collagen 1α1b of Sebastes umbrosus. |
| Haddock (IV)              | GVTGSPGpGDPGK       | Gadiformes (taxid 8043), hit with collagen 1α1-like of Gadus morhua (XP_030195580.1). |
| Salmon (II)               | GSTGAAGISGAPGpGTR   | Percifomes (taxid 8111), no hits.       |
| Sea bass (II)             | GNGDH1GApGPK        | See trout.                              |
| Pangasius (II)            | GSpGPAITGpGpGTR     | Siluriformes (taxid 7995), hit with collagen 1α1b of Pangasianodon hypophthalmus. |
| Tilapia (I)               | GAgpAIGGAGpGpGAR    | Gichifomes (taxid 1489911), hit with collagen 1α1b of Neolamprologus brichardi, Oreochromis aureus, Oreochromis niloticus and with collagen 1α1-like of Archocentrus unrochus. |

Table 3
Summary of the highest sequence coverages obtained for the analyzed fish samples, searched against the collagen 1α3 fasta file (see Supplementary Table 1).

| Animal species (product) | Order   | Family    | Highest sequence coverage | Species with highest sequence coverage   |
|--------------------------|---------|-----------|---------------------------|----------------------------------------|
| Barramundi               | Percifomes | Latidae  | 33.0 | Toxotes jaculatus |
| Sea bream                | Percifomes | Sparidae | 47.8 | Acanthopagrus latu |
| Trout                    | Salmoniformes | Salmonidae  | 59.5 | Oncorhynchus mykiss |
| Sardine                  | Clupeiformes | Clupeidae  | 11.2 | Carassius auratus |
| Hake                     | Gadiformes | Gadidae  | 14.6 | Amphilipon occelaris |
| Cod                      | Gadiformes | Gadidae  | 15.9 | Nothobranchus kuhntae 2 |
| Saithe                   | Gadiformes | Gadidae  | 15.6 | Nothobranchus kuhntae 2 |
| Mackerel                 | Percifomes | Scrombriadae  | 17.7 | Micrurypress salmoide |
| Rose fish                | Scorpifomes | Sebastidae  | 47.5 | Sebastes umbrosus |
| Haddock                  | Gadiformes | Gadidae  | 12.7 | Nothobranchus kuhntae 2 |
| Salmon                   | Salmoniformes | Salmonidae  | 42.2 | Oncorhynchus mykiss |
| Sea bass                 | Percifomes | Moronidae | 48.2 | Morone saxatilis |
| Pangasius                | Siluriformes | Pangasidae  | 62.8 | Pangasianodon hypophthalmus |
| Tilapia                  | Cichifomes | Cichidae  | 61.2 | Oreochromis niloticus |
| Cod                      | Gadiformes | Gadidae  | 13.8 | Nothobranchus kuhntae 2 |
| Salmon                   | Salmoniformes | Salmonidae  | 36.7 | Salvelinus namaycush |

The accession title were considered, it was decided to use the decision of collagen 1α1-like sequences for the category IV species. To facilitate identification for the Gadiformes hake, cod, saithe and haddock two different Gadus morhua collagen 1α1-like sequences were used: XM_03039720.1 (mRNA sequence) and XP_030196341.1 (protein). Clustal Omega analysis indicated that XM_03039720.1 clusters with collagen 1α3 and XP_030196341.1 with collagen 1α1, see Supplementary Fig. 1. Collagen 1α1-like is therefore not a synonym of collagen 1α3 and 1α1b. In Supplementary Fig. 3, an MS/MS spectrum is presented from a collagen 1α1-like peptide from accession XM_03039720.1 occurring in the cod gelatin sample. The identified peptide GPA-GAQGGLGAPGPK differs from the analogus part in XP_030196341.1 by 4 amino acids. The elucidated cod peptide sequence only provided a hit with (translated) XM_03039720.1 when searched against taxid 8043 (Gadiformes) using protein blast. Therefore, it is probable that collagen 1α3 was identified in cod. The peptide GPA-GAQGGLGAPGPK was also detected in a deamidated form, at residue Q6. Accession XM_03039720.1 could be used to find specific peptides in the other Gadiformes hake, saithe and haddock, see Table 2. Similarly, the Clupea harengus collagen 1α1-like sequence XP_031416510.1 (only protein) clustered with collagen 1α3, see Supplementary Fig. 1, and could be used to extract information for sardine. Using this approach a specific peptide for sardine could be identified and therefore it is probable that collagen 1α3 was identified in sardine, see Table 2. However, accurately annotated genetic data is required to fully confirm these results, especially for category IV species.

Collagen 1α3 could be identified in several teleost fish species, with varying degrees of confidence. Depending on the amount of available genetic information it was important to use different approaches to identify the collagen chain. It is expected that collagen 1α3 occurs in many other teleost fish species. Additionally, collagen-specific collision induced internal fragmentations were described. During the interpretation of MS/MS spectra it is very important to consider these fragmentations to facilitate elucidation of the sequence and to avoid wrong protein and peptide assignments. Collagen 1α3 was identified not only in fish skin extracts, but also in fish skin gelatin which is a food product. Its high abundance, having the same stoichiometry as the 1α1 and 1α2 chains in fish skin, underlines that the protein is more important for food consumers than suggested by its prominence in literature. Although collagen 1α3 is similar to collagen 1α1, its presence will affect food (product) properties, such as amino acid content. Tilapia collagen 1α3, for example, contains significantly less methionine in the chain than tilapia collagen 1α1 and significantly more histidine, serine and isoleucine. This will also affect the population of collagen di- and tripeptides which survive the gastrointestinal tract and brush border.
peptidase activity, to finally enter the blood (Kleinnijenhuis et al., 2020). Due to evolutionary divergence, however, the effect on amino acid content will be different for each fish species.

Conclusion

Different approaches were used to identify collagen Iα3 in the skin of teleost fish species. Depending on the amount of available genetic information the chain could be identified with varying degrees of confidence. Due to the specific collagen properties and behavior in LC-MS/MS it is essential to also manually interpret MS/MS spectra of tryptic collagen peptides, to avoid wrong protein and peptide assignments. We described the occurrence of collagen-specific internal y(3n-2)/b and y(3n)/a fragment ion series, especially abundant when the third residue is a hydroxyproline (Hyp-3 peptides). Collagen Iα3 was identified in teleost fish skin extracts, but also in fish skin gelatin which is a food product. Although it is similar to collagen 1α1, the presence of collagen Iα3 will affect food (product) properties, such as amino acid content and the population of collagen di- and tripeptides which can enter the blood. Therefore, it is important to consider the presence of collagen Iα3 in fish (products) and nutraceuticals prepared from fish skin collagen.

CRediT authorship contribution statement

Anne J. Kleinnijenhuis: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Frédérique L. van Holthoon: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Bastiaan van der Steen: Conceptualization, Resources, Writing – review & editing.

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.

Acknowledgement

The work was financed by Triskelion.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2022.100333.

References

Bellamy, G., & Bornstein, P. (1971). Evidence for procollagen, a biosynthetic precursor of collagen. Proceedings of the National Academy of Sciences USA, 68, 1138–1142.

Boyd, P. R., Tuckwell, D. S., Plumb, D. A., Rock, C. F., & Poulsom, R. (2003). A novel and highly conserved collagen (proα1(XVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. Journal of Biological Chemistry, 278, 31067–31077.

Brie, L., Tabb, D. L., Yates, J. R., & Wysoczki, V. K. (2003). Clearage N-terminal to proline: Analysis of a database of peptide tandem mass spectra. Analytical Chemistry, 75, 1963–1971.

Carlton, P., Smalley, D. M., & Van Beneden, R. J. (2013). Prototoxic analysis of arsenic-exposed zebrafish (Danio rerio) identifies altered expression in proteins involved in fibroblast and lipid uptake in a gender-specific manner. Toxicological Sciences, 134, 83–91.

Duconseille, A., Astruc, T., Quintana, N., Meursman, F., & Sante-Lhoutellier, V. (2015). Gelatin structure and composition linked to hard capsule dissolution: A review. Food Hydrocolloids, 43, 360–376.

Exposito, J. Y., Cluzel, G., Garrone, B., & Letihan, C. (2002). Evolution of collagens. The Anatomical Record, 268, 302–316.

Exposito, J. Y., Larroux, C., Cluzel, G., Valcourt, U., Letihan, C., & Degnan, B. M. (2008). Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. Journal of Biological Chemistry, 283, 28226–28235.

Exposito, J. Y., Valcourt, U., Cluzel, C., & Letihan, C. (2006). The fibrillar collagen family. International Journal of Molecular Sciences, 11, 407–426.

Fratzl, P. (2008). Collagens: Structure and mechanics, an introduction. Boston, MA: Springer.

Gudmundsson, M., & Hafsteinsson, H. (1997). Gelatin from cod skins as affected by chemical treatments. Journal of Food Science, 62, 37–39.

Karsdal, M. A., Leeming, D. J., Henriksen, K., & Bay-Jensen, A. (2016). Biochemistry of collagens, laminins and elastin. Structure, function and biomarkers. Elsevier Inc.

Leggatt, R. A., Jefferson, A., Koth, J., Wang, X., Killen, A. C., Stockdale, W. T., Potts, H. G., ... & Naylor, B. B. (2013). Optimization of fish gelatin extraction from skins and bones: A comparative study. Bioanalytical Chemistry, 412, 973–992.

Kim, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., ... & Rohksar, D. (2008). The genome of the chononflagellate Monosiga brevicollis and the origin of metazoans. Nature, 451, 783–788.

Kleinnijenhuis, A. J., Van Holthoon, F. L., & Herregroes, G. (2018). Validation and theoretical justification of an LC-MS method for the animal species specific detection of gelatin. Food Chemistry, 243, 461–467.

Kleinnijenhuis, A. J. (2019). Visualization of genetic drift processes using the conserved collagen 1α1 GXY domain. Journal of Molecular Evolution, 87, 106–136.

Kleinnijenhuis, A. J., Van Holthoon, F. L., Maathuis, A. J. H., Vanhoecke, B., Prawitt, J., Wasquier, F., & Wittrant, Y. (2020). Non-targeted and targeted analysis of collagen hydrolysates during the course of digestion and absorption. Analytical and Bioanalytical Chemistry, 412, 973–992.

Koli, J. M., Basu, S., Guipidip, V., Chouksay, M. K., & Nayak, B. B. (2013). Optimization of fish gelatin extraction from skins and bones: A comparative study. Ecology, Environment and Conservation, 19, 47–56.

Koth, J., Wang, X., Kilien, A. C., Stockdale, W. T., Potts, G. H., Jefferson, A., ... & Naylor, B. B. (2010). Runx1 promotes scar deposition and inhibits myocardial proliferation and survival during zebrafish heart regeneration. Development, 137, dev185660.

Leggatt, R. A., & Iwama, G. K. (2003). Occurrence of polyploidy in the fishes. Reviews in Fish Biology and Fisheries, 13, 237–246.

Makareeva, E., & Lekin, S. (2014). Collagen structure, folding and function. In: Osteogenesis imperfecta. Academic Press, Cambridge, MA. ISBN: 978-0-12-397165-4.

Morvan-Dubois, G., Le Guellec, D., Garrone, R., Zylberberg, L., & Bonnau, L. (2003). Phylogenetic analysis of vertebrate fibrillar collagen locates the position of zebracid α2(10) and suggests an evolutionary link between collagen α chains and Hox clusters. Journal of Molecular Evolution, 57, 501–514.

Piez, K. A. (1965). Characterization of a collagen from codfish skin containing three structurally different α chains. Biochemistry, 4, 2590–2596.

Rutnagar, B. (1985). Collagen gene construction and evolution. Journal of Molecular Evolution, 22, 141–149.

Shahidi, F. (2007). Maximising the value of marine by-products. Woodhead Publishing Ltd.

Shiendries, M. D., & Raines, R. T. (2009). Collagen structure and stability. Annual Review of Biochemical, 78, 929–958.

Sicot, F. X., Exposito, J. Y., Masselot, M., Garrone, R., Deutsch, J., & Gaill, F. (1997). Cloning of an annelid fibrillar-collagen gene and phylogenetic analysis of vertebrate and invertebrate collagens. European Journal of Biochemistry, 246, 50–58.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7, 539.

Stevens, P. (2010). Gelatine. In: Food stabilisers, thickeners and gelling agents. Blackwell Publishing Ltd., Chichester, United Kingdom, 116–144. ISBN: 978-1-444-31472-4.

Teuchan, U., & Scholz, C. B. (2003). Origin and evolution of endoderm and mesoderm. International Journal of Developmental Biology, 47, 531–539.

Vergauwen, B., Stevens, P., Prawitt, J., Ollive, J., Bruns, E., & iBael, W. (2016). Gelatin. Ullman’s Encyclopedia of Industrial Chemistry. Wiley-VCH, 1–22.

Yang, H., Butler, E. R., Monier, S. A., Teubl, J., Fenyo, D., Ueberheide, B., & Siegel, D. (2021). A predictive model for vertebrate bone identification from collagen using proteomic mass spectrometry. Scientific Reports, 11, 10990.

Yasu, N., Ono, K., Konomi, H., & Nagai, Y. (1984). Transitions in collagen types during endochondral ossification in human growth cartilage. Clinical Orthopaedics and Related Research, 183, 215–218.

Zhang, X., Boot-Hendrickx, R. P., Huxley-Jones, J., Forse, L. M., Nould, A. P., Robertson, D. L., ... & Sarras, M. P., Jr. (2007). The collagen of hydra provide insight into the evolution of metazoan extracellular matrices. Journal of Biological Chemistry, 282, 6792–6802.