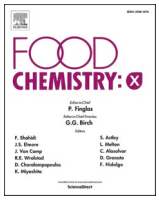
Food Chemistry: X 14 (2022) 100333

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



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

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

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

ARTICLE INFO

Keywords:
Collagen 1α3
Teleost
Fish
LC-MS
Gelatin
Internal fragmentation

ABSTRACT

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). Choanoflagellate 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 eumetazoan 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, & Poulsom, 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 (Makareeva, & 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 1α2 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 Guellec, Garrone, Zylberberg, & Bonnaud, 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

* Corresponding author.
E-mail address: anne.kleinnijenhuis@triskelion.nl (A.J. Kleinnijenhuis).

https://doi.org/10.1016/j.fochx.2022.100333
Received 3 August 2021; Received in revised form 22 March 2022; Accepted 15 May 2022
Available online 17 May 2022

Table 1
Fibril-forming collagen types 1, 2, 3, 5, 11, 24 and 27 with their corresponding clades and typical presence in important tissues. A larger overview of collagen types and characteristics has been extensively described elsewhere (Duconseille, Astruc, Quintana, Meersman, & Sante-Lhoutellier, 2015).

Type	Clade	Typical presence in
1	A	Bone, skin, tendon
2	A	Cartilage
3	A	Skin
5	A (5α2), B (5α1, 5α3)	Bone, skin
11	B	Cartilage
24	C	Bone
27	C	Cartilage

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 polyploidy 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 Acquity 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 method. The mass range was set to m/z 200–2000 at a resolution of 35,000. 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

allowed up to one missed cleavage; the precursor mass tolerance was set at 10 ppm, and fragment mass tolerance at 0.02 Da. Carbamidomethylation (C) was set as a fixed modification, although cysteine is not expected to occur in collagen. Oxidation (M, P and K) were set as variable modifications. False discovery rate on the peptide level was less than 1%. Only peptides labeled 'high confidence' were considered (q-value < 0.01).

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 1α3/1α1b 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 5α2 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 1α3 and 1α1b sequences versus the relations with the collagen 1α1 sequences indicate that the chain names 1α3 and 1α1b 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 sequences) 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 1α3. Additionally, the Clustal Omega analysis revealed that the selected collagen 1α1 sequences formed an individual cluster within the set of investigated sequences.
b) The similarities between the collagen 1α3 and 1α1 sequences indicate that collagen 1α3 did not emerge long before teleost radiation, confirming earlier findings (Kimura, Ohno, Miyauchi, & Uchida, 1987). Therefore, it is difficult to discriminate between collagen 1α1 and 1α3 in fish species when there is limited genetic or protein sequence information available.
c) On the cDNA level, collagen 1α3 and 1α1 are most similar to collagen 2α1, followed by 1α2 and 3α1.
d) The added cod (*Gadus morhua*) collagen 1α1-like sequence XM_030339720.1, which will be discussed in more detail below, is more similar to the collagen 1α3 sequences than to the collagen 1α1 sequences. It is probable that this cod collagen 1α1-like sequence thus represents collagen 1α3, which makes it reasonable to explore the feasibility of using collagen 1α1-like sequences in the search for collagen 1α3.

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 1α3 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 1α3. To assign collagen 1α3, 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

Species		Bos taurus	Sus scrofa	Gallus gallus	Latimeria chalumnae	Oreochromis niloticus	Bos taurus	Sus scrofa	Gallus gallus	Latimeria chalumnae	Oreochromis niloticus	Anguilla japonica	Pygocentrus nattereri	Carassius auratus	Esox lucius	Gadus morhua	Acanthopagrus latus	Morone saxatilis	Sebastes umbrosus	Pangasianodon hypophthalmus	Salvelinus namaycush	Oncorhynchus keta	Oncorhynchus mykiss	Oreochromis niloticus	Bos taurus	Sus scrofa	Gallus gallus	Latimeria chalumnae	Oreochromis niloticus	Bos taurus	Sus scrofa	Gallus gallus	Latimeria chalumnae
	Type	1α1	1α1	1α1	1α1	1α1	1α2	1α2	1α2	1α2	1α2	1α3	1α1b	1α3	1α1b	1α1 like	1α1b	1α1b	1α1b	1α1b	1α1b	1α1b	1α3	1α3	2α1	2α1	2α1	2α1	2α1	3α1	3α1	3α1	3α1
Bos taurus	1α1	0	135	604	645	768	1024	1012	992	1034	1025	715	823	806	884	939	760	777	858	832	895	897	894	785	864	858	894	819	876	1077	1057	1130	1088
Sus scrofa	1α1	135	0	591	619	736	1028	1010	983	1028	1009	699	813	792	863	915	743	755	845	803	878	874	871	772	844	856	873	789	868	1077	1052	1113	1064
Gallus gallus	1α1	604	591	0	764	845	1121	1121	1098	1108	1096	817	896	908	945	890	845	884	923	935	962	961	955	875	930	931	850	921	981	1155	1146	1205	1167
Latimeria chalumnae	1α1	645	619	764	0	753	986	958	900	942	1005	695	775	766	857	943	753	744	820	770	879	869	862	735	859	868	926	751	878	1085	1051	1090	1029
Oreochromis niloticus	1α1	768	736	845	753	0	1069	1047	1013	1063	944	678	677	730	853	825	653	633	721	711	849	844	845	665	888	891	936	873	849	1127	1111	1144	1133
Bos taurus	1α2	1024	1028	1121	986	1069	0	236	645	932	1001	1044	1064	1051	1135	1210	1061	1051	1107	1074	1142	1129	1130	1036	1035	1055	1124	1028	1082	1220	1235	1255	1223
Sus scrofa	1α2	1012	1010	1121	958	1047	236	0	598	893	972	1017	1035	1043	1114	1207	1032	1020	1081	1019	1127	1119	1117	1007	1034	1023	1134	992	1037	1210	1203	1225	1191
Gallus gallus	1α2	992	983	1098	900	1013	645	598	0	837	975	988	1014	1030	1118	1195	1022	1024	1078	1017	1135	1125	1124	1008	1040	1019	1115	944	1023	1183	1181	1215	1166
Latimeria chalumnae	1α2	1034	1028	1108	942	1063	932	893	837	0	973	1041	1045	1029	1146	1222	1036	1052	1075	1021	1146	1139	1145	1013	1068	1067	1140	959	1075	1194	1205	1205	1147
Oreochromis niloticus	1α2	1025	1009	1096	1005	944	1001	972	975	973	0	992	962	972	1071	1117	949	955	1015	977	1092	1083	1087	942	1058	1040	1113	1031	1011	1214	1207	1227	1225
Anguilla japonica	1α3	715	699	817	695	678	1044	1017	988	1041	992	0	644	650	740	781	616	601	669	653	730	719	718	602	891	895	939	893	892	1144	1135	1178	1118
Pygocentrus nattereri	1α1b	823	813	896	775	677	1064	1035	1014	1045	962	644	0	516	761	820	642	625	691	447	768	764	762	631	925	913	1000	897	879	1116	1099	1180	1124
Carassius auratus	1α3	806	792	908	766	730	1051	1043	1030	1029	972	650	516	0	796	832	640	633	697	570	800	784	780	625	924	925	966	870	900	1118	1114	1177	1146
Esox lucius	1α1b	884	863	945	857	853	1135	1114	1118	1146	1071	740	761	796	0	827	711	721	773	779	411	417	419	738	1040	1058	1038	1011	1020	1159	1150	1223	1176
Gadus morhua	1α1 like	939	915	890	943	825	1210	1207	1195	1222	1117	781	820	832	827	0	690	683	684	863	815	810	802	736	1033	1048	1057	1082	1027	1296	1290	1313	1314
Acanthopagrus latus	1α1b	760	743	845	753	653	1061	1032	1022	1036	949	616	642	640	711	690	0	261	391	671	715	710	710	365	933	923	984	910	849	1151	1137	1185	1169
Morone saxatilis	1α1b	777	755	884	744	633	1051	1020	1024	1052	955	601	625	633	721	683	261	0	365	643	729	723	722	326	923	917	975	894	838	1152	1144	1185	1148
Sebastes umbrosus	1α1b	858	845	923	820	721	1107	1081	1078	1075	1015	669	691	697	773	684	391	365	0	745	760	753	754	448	985	970	1008	962	906	1210	1197	1232	1224
Pangasianodon hypophthalmus	1α1b	832	803	935	770	711	1074	1019	1017	1021	977	653	447	570	779	863	671	643	745	0	795	786	787	639	934	935	1009	892	894	1124	1120	1154	1114
Salvelinus namaycush	1α1b	895	878	962	879	849	1142	1127	1135	1146	1092	730	768	800	411	815	715	729	760	795	0	64	67	760	1034	1032	1060	1005	1012	1195	1194	1238	1195
Oncorhynchus keta	1α1b	897	874	961	869	844	1129	1119	1125	1139	1083	719	764	784	417	810	710	723	753	786	64	0	33	752	1037	1023	1060	994	1006	1203	1192	1232	1192
Oncorhynchus mykiss	1α3	894	871	955	862	845	1130	1117	1124	1145	1087	718	762	780	419	802	710	722	754	787	67	33	0	751	1035	1028	1053	996	1009	1200	1198	1237	1191
Oreochromis niloticus	1α3	785	772	875	735	665	1036	1007	1008	1013	942	602	631	625	738	736	365	326	448	639	760	752	751	0	944	932	989	899	853	1135	1116	1164	1133
Bos taurus	2α1	864	844	930	859	888	1035	1034	1040	1068	1058	891	925	924	1040	1033	933	923	985	934	1034	1037	1035	944	0	186	625	620	693	1104	1101	1168	1158
Sus scrofa	2α1	858	856	931	868	891	1055	1023	1019	1067	1040	895	913	925	1058	1048	923	917	970	935	1032	1023	1028	932	186	0	646	618	685	1119	1110	1182	1150
Gallus gallus	2α1	894	873	850	926	936	1124	1134	1115	1140	1113	939	1000	966	1038	1057	984	975	1008	1009	1060	1060	1053	989	625	646	0	679	753	1146	1141	1197	1191
Latimeria chalumnae	2α1	819	789	921	751	873	1028	992	944	959	1031	893	897	870	1011	1082	910	894	962	892	1005	994	996	899	620	618	679	0	639	1089	1086	1100	1070
Oreochromis niloticus	2α1	876	868	981	878	849	1082	1037	1023	1075	1011	892	879	900	1020	1027	849	838	906	894	1012	1006	1009	853	693	685	753	639	0	1136	1119	1146	1129
Bos taurus	3α1	1077	1077	1155	1085	1127	1220	1210	1183	1194	1214	1144	1116	1118	1159	1296	1151	1152	1210	1124	1195	1203	1200	1135	1104	1119	1146	1089	1136	0	266	909	1038
Sus scrofa	3α1	1057	1052	1146	1051	1111	1235	1203	1181	1205	1207	1135	1099	1114	1150	1290	1137	1144	1197	1120	1194	1192	1198	1116	1101	1110	1141	1086	1119	266	0	892	1018
Gallus gallus	3α1	1130	1113	1205	1090	1144	1255	1225	1215	1205	1227	1178	1180	1177	1223	1313	1185	1185	1232	1154	1238	1232	1237	1164	1168	1182	1197	1100	1146	909	892	0	960
Latimeria chalumnae	3α1	1088	1064	1167	1029	1133	1223	1191	1166	1147	1225	1118	1124	1146	1176	1314	1169	1148	1224	1114	1195	1192	1191	1133	1158	1150	1191	1070	1129	1038	1018	960	0

Fig. 1. Distance table of 32 compared animal species and collagen types, regarding the collagen GXY domain, indicating, from green via yellow to red, the increasing amount of nucleotide differences.

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 collagen properties and behavior in LC-MS/MS, as outlined above, should be taken into account to reduce the risk of false-positive identifications.

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 in bold in Fig. 2 and represents residues 220–237 of the GXY domain. A combined extracted 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 position and therefore many tryptic collagen peptides have an N-terminal glycine residue. Hyp residues are very often located in the third GXY position. As a result, a large part of the tryptic collagen peptides has Hyp in the third position from the N-terminus, which is also the case for the tilapia collagen 1α1 and 1α3 peptides shown in Fig. 4. In the remainder of the document we designate these peptides as Hyp-3 peptides. When Hyp-3 peptides with n residues are fragmented using CID an internal $y_{(n-2)}$/b fragment ion series is produced starting with the **p**G$^+$ ions at *m/z* 171.076 (Kleinnijenhuis, Van Holthoon, & Herregods, 2018). For the tilapia collagen 1α1 peptide the series extends to **p**GAAGVAGA$^+$ and for the tilapia collagen 1α3 peptide to **p**GAAGIAGA$^+$. Regularly, $y_{(n-2)}$/a-type internal fragments, which are 27.995 Da (–CO) lower in mass, can be observed as well, but usually with lower intensity than the $y_{(n-2)}$/b internal fragments. Finally, an extensive y ion series is typically present up to $y_{(n-2)}$. 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 **p**G$^+$, 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 **p**G$^+$.

For several other analyzed fish species, it was not possible to find the tryptic collagen 1α3 peptide 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

Collagen type	Aligned collagen helix sequence	position
1α1	GPMGPMGPRGPPGPPGSSGPQGFTGPPGEAGEPGSPGPMGPRGPAGPPGKNGEDGESGKPGRPGERGPPGPQGARGFPGTPGLPGIKGHRGFSGLDGAKGDTGPAGPKGEAGTPGENGTPGAMGPRGLPGERGRAGATGAAGARGNDGAA	150
1α2	GPMGLMGPRGPPGPPGASGPQGHTGHAGEPGEPGQAGAVGPRGPPGPPGKAGEDGNNGKPGKPGDRGAPGPQGARGFPGTPGLPGMKGHRGYTGLDGRKGEPGAAGPKGEPGAHGAAGSPGLAGARGLPGPRGERGPAGPAGAKGADGNA	150
1α3	GPIGPMGPRGPPGPSGPPGPQGFTGPPGEPGEPGASGPLGSKGPSGPPGKNGDDGEAGKPGKPGERGAPGPQGGKGFPGTPGLPGIKGHRGFSGLDGAKGDSGPAGPKGEPGTSGENGIPGALGARGLPGERGRPGAPGPAGARGNDGNT	150
2α1	GPMGPMGPRGPPGPTGAPGPQGFQGNPGEAGEPGQSGPMGPRGPPGPPGKPGDDGEAGKPGKSGERGPSGPQGARGFPGTPGLPGIKGHRGYPGNDGARGETGAVGAKGESGSPGESGAPGPMGPRGLPGERGRPGASGAAGARGNDGLP	150
1α1	GAAGPPGPTGPAGPPGFPGGPGAKGDAGQGGKPGEPGPAGAKGEPGNPGPAGPAGPAGNPGSDGAPGAKGAPGPFGPGPRGSPGQGAAGAPGPKGNTGEAGAPGSKGEAGAKGEAGAPGVQGPPGPPGEEGKRGAKGEPGAA	300
1α2	GPTGPAGPLGAAGPPGFPGGPGPKGETGPVGATGPSGPQGSKGEPGPNGAVGFVGPSGNPGANGLNGAKGAAGTPGVAGAPGFPGPRGGPGPQGAAGPKGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGDK	300
1α3	GPAGPPGPTGPAGPPGFPGGAGAKGETGPAGKGSEGPQGAKGEPGNPGPAGPAGPAGAPGSDGSPGSKGAPGAAGIAGAPGFPGARGPPGPQGPSGPKGEPGAKGEPGPAGIQGLPGPSGEEGKRGGKGEPGGA	300
2α1	GPAGPPGPVGPSGAPGFPGSPGAKGEAGPTGAKGPEGAQGPRGESGTPGSPGPSGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPKGPPGPQGATGPLGPKGTSGDPGIPGFKGEAGPKGEIGPPGLQGPNGPQGEEGKRGPKGEPGEPGSA	300
1α1	GARGGPGERGAPGGRGFPGSDGPAGPKGATGEPGKTGEPGLPGAKGMTGSPGNPGPDGKIGPSGAPGQDGRPGPPGPGGAKGQPGVMGFPGPKGAAGEAGKPGERGTMGPTGPAGAPGKDGDVGAQGPPGPAGPA	450
1α2	GLAGAKGAIGAPGSKGMPGAEGTGPVGMPGAKGATGAAGPIGPPGDAGIAGEPGAAGLKGLPGSPGSSGPPGKEGPAGPSGQDGSKGPPGPSGPKGLSGNIGFPGPKGPSGEPGKPGEKGATGPTGLKGPPGPDGNNGATGATGVAGGP	450
1α3	GPKGPPGERGAPGAKGFPGADGGAGKGAPGERGAPGPLGAQGVTGESGSPGAPGAPGLKGVTGSPGSPGPDGKTGAPGAPGAPGQDGRPGPPGPGGSKGEQGPSGQPGIMGFPGPKGASGDSGKPGERGATGPAGAVGAPGPKGNNGDPGPSGPKGEPGAKGEPGPAGIQGLPGPSGEEGKRGGKGEPGGA	450
2α1	GPKGPPGERGAPGNKGFPGQDGLAGPKGAPGERGPSGASGPKGASGDPGTPGEPGLPGAKGLTGRPGDAGPQGKVGPSGAPGEDGKPGPPGPQGAKGQPGVMGFPGPKGATGEPGKSGEKGLAGAPGLKGLPGKDGETGSAGPPGPAGPA	450
1α1	GERGEQGPAGSPGFQGLPGPQGAVGETGKPGEQGVPGEAGAPGPAGAKGDKGPRGATGIPGAPGAIGPAGAKGSPGASGNDGAKGDAGAPGTPGAQGPPGLQGMPGERGAAGLPGLKGNKGDQGPKGADGTPGKDGPRGLTGPIGLPGPAGSP	600
1α2	GEKGEQGPSGSPGFQGLPGPAGPTGEAGKPGDQGIPGEPGAAGNAGAPGEKGNPGAAGSAGPQGPIGPRGPAGPGDGGKGEPGPAGVAGAPGHQGAGGMPGERGGAGTPGPKGEKGEPGHKGPDGNPGTDGPRGLAGPAGPPGPTGAN	600
1α3	GEKGEQGPAGSPGFQGLPGPAGPTGEAGKPGDQGVPGEAGEAGPSGPAGPTGIAGPTGPRGAPGAGNDGPTGPEPGAAGAPGGLGAPGMQGMPGERGAAGLPGAKGDNGEPGGKGVDGAPGKDGPRGLTGPIGVPGPPGAQ	600
2α1	GERGEQGQPGPSGFQGLPGPPGPPGEGGKPGDQGVPGEAGPAGATGPRGEKGFPGERGAAGSQGLQGPRGLPGTPGTDGPKGAIGPAGTAGAQGPPGLQGMPGERGGAGIPGPKGDKGDVGEKGPEGAPGKDGARGLTGPIGPPGPSGPN	600
1α1	GDRGEPGAQGPVGPSGAKGPPGEKGEAGPPGPAGFAGPPGADGQPGAKGEPGDNGAKGDSGPPGPAGPTGAPGPQGPVGNTGPKGANGPAGPPGATGFPGAAGRVGPPGPAGNAGPPGPPGPAGKEGPKGNKGETGPAGPPGELGAAGPP	750
1α2	GDKGEQGSFGPAGPAGPKGPSGEKGEVGPAGAPGFAGPPGADGQAGAKGEKGPSGAKGEVGPSGLAGPAQQSGPAGPAGPGGPPGANGDNGPPGLTGFPGAAGRVGAAGPAGIVGPPGPAGPSGKDGPKGPKGDPGSPSGEPGIIGPP	750
1α3	GERGEAGAVGVAGPSGPKGSPGDKGEPGPAGAAGFAGPPGVDGQPGAKGESGDSGPKGDAGPPGPTGPAGASGPQGPAGAPGPKGANGSVGSPGATGFPGPAGRVGPPGPAGAGGAPGPGGPTGKDGADGKGEKGETGPAGPPGEAGAAGPP	750
2α1	GERGESGPAGPSGAPGVKGAPGDKGETGPPGPAGFAGPPGADGQPGKPGEQGEPGQKGDAGAPGPQGPSGAPGPAGPTGVSGPKGANGAQGPPGATGFPGAAGRVGPPGPNGNPGPPGAPGPPGKDGPRGKVGSDGGPPGKQGDAGLKGPA	750
1α1	GPPGEKGSPGADGAPGSAGIPGPQGIAGQRGIVGLPGQRGERGFPGLAGPVGEPGKQGPSGPSGERGPPGPMGPPGLAGAPGEPGEQGPKGADGPRGPDGPKGVVGPAGKTGDKGETGPAGPAGAA	900
1α2	GLAGEKGPSGESGPPGPSGAPGTSGPLGLQGFVGLPGSKGDRGAPGGAGGVGEPGNLGPAGPPGANGAPGNIGLPGMTGPQGEAGIEGSPGNDGPPGTPGAAGLKGDKGEPGSAGTTGLAGAPGPAGPTGAAGERGEMGFAGIKGETGLAGPPGPPGAPGGVGPSGKTGDKGESGPSGPAGPA	900
1α3	GPSGEKGSAGSDGAPGAPGIPGPQGIAGQRGIVGLPGQRGERGFPGLPGPSGEPGKQGSSGPVGERGAPGPAGPPGLSGPTGEAGREGSPGHDGAPGKDGPSGPKGDKGETGLAGPPGPAGPAGPPGAPGGVGPSGKTGDKGESGPSGPAGPA	900
2α1	GASGEKGDAGEDGPPGPPGPSGPQGLAGQRGIVGLPGQRGERGFPGLPGPSGEPGKQGASGGPGDKGPPGPVGPPGLTGPAGEPGIEGTPGSDGPPGKDGSPGHKGERGNTGPAGAPGAPGAPGANGPVGPTGKQGDKGEPGPQGPAGSS	900
1α1	GPAGPKGPAGAPGLAGDKGETGEAGERGMKGHRGFTGMQGPPGPPGTSGESGPAGAAGPAGPKGPSGAAGAPGKDGVSGLPGPTGPPGPRGHSGEMGPAGPPGPPGAPGPPGAPGAP	1014
1α2	GPAGANGASGPAGPMGERGVAGDKGEIGMPGLKGHPGLQGMPGPSGPPGDTGAAGAHGPSGPKGPAGPHGPVGKDGRNFGAHGTMGAPGAKGPNGYSGPVGPPGPPGLPGPPGPA	1014
1α3	GPAGVKGPAGPAGAKGDKGEAGEAGEKGHKGHRGFTGASGLPGPAGPPGERGPAGASGPAGPKGPSGSNGAPGKDGVNGIPGPIGPPGPRGNGEMGPAGPPGPPGAPGPPGPPGP	1014
2α1	GPAGANGMPGPQGPKGDKGEAGEAGEKGQKGHRGFTGLQGLPGPPGQPGDQGATGPAGPAGQKGPPGPVGPAGKDGTNGLPGPIGPPGPRGNSGETGPAGPPGPNPGPPGPPGPP	1014

Fig. 2. Aligned GXY domain sequences of tilapia (*Oreochromis niloticus*) collagen 1α1, 1α2, 1α3 and 2α1. Trypsin 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.

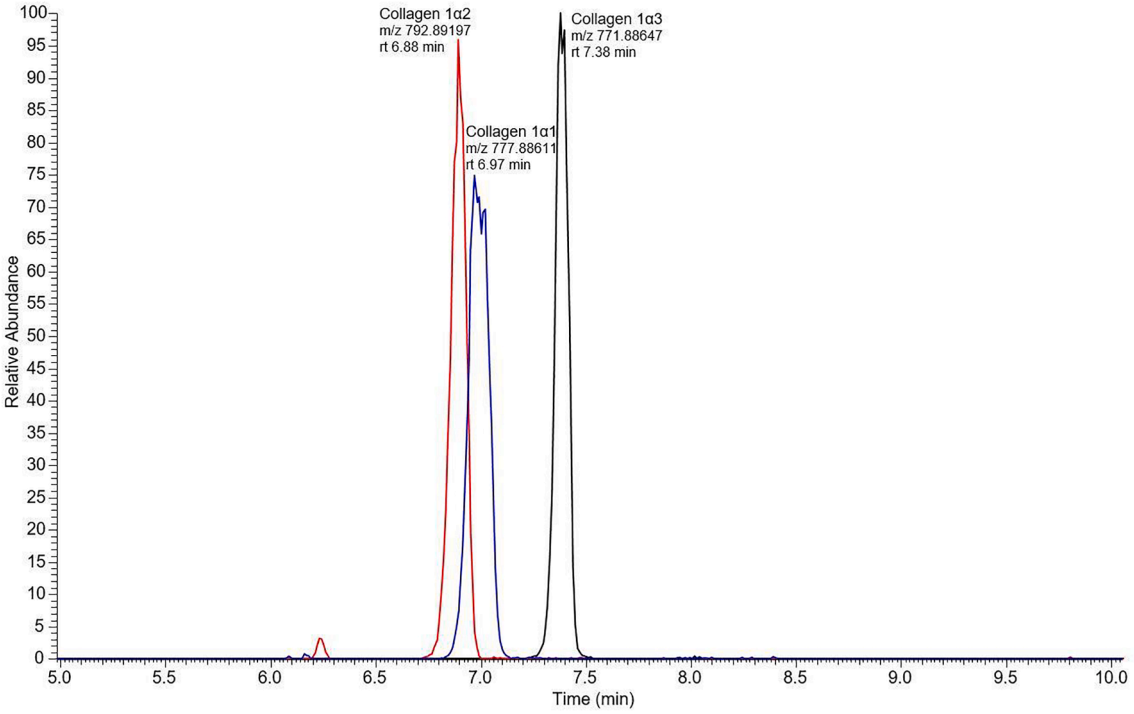


Fig. 3. Combined extracted chromatogram of analogous peptides from tilapia collagen 1α1 (GA**p**GAAGVAGA**p**GF**p**GPR), collagen 1α2 (GAAGT**p**GVAGA**p**GF**p**GPR) and collagen 1α3 (GA**p**GAAGIAGA**p**GF**p**GAR), present in tilapia skin gelatin.

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, of tryptic 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 GSTGAAGISGA**p**GF**p**GTR 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

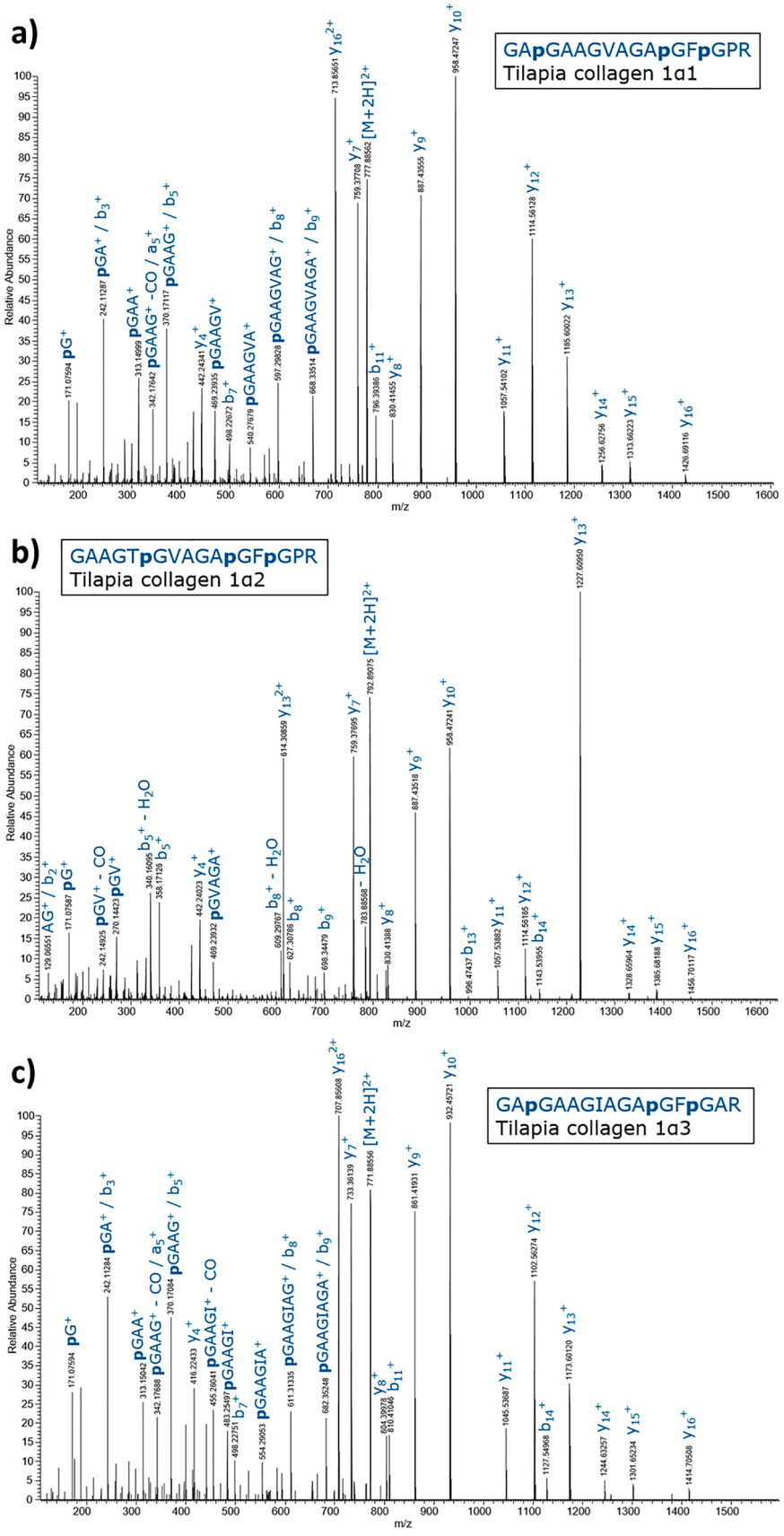


Fig. 4. MS/MS spectra of tryptic tilapia collagen 1 peptides from tilapia skin gelatin; a) 1α1, b) 1α2 and c) 1α3.

Table 2
Manually confirmed collagen 1α3 peptide identifications. In the third column the protein blast results are summarized.

Animal species (category)	Elucidated peptide	Searched against, description 100% hits
Barramundi (III)	GPAGAQGGVGA**p**GPK	Perciformes (taxid 8111), no hits.
Sea bream (II)	EGSQGHDGA**p**GR	Perciformes (taxid 8111), no hits. Hits with collagen 1α1b of *Acanthopagrus latus* and with collagen 1α1-like and 2α1-like of *Sparus aurata* when searched against Sparidae (taxid 8169).
Trout (II)	GSTGAAGISGA**p**GF**p**GTR	Salmoniformes (taxid 8006), hit with collagen 1α1(b)(-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)	GATGS**p**GIAGA**p**GF**p**GPR	Clupeiformes (taxid 32446), hit with collagen 1α1-like of *Clupea harengus* (XP_031416510.1).
Hake (IV)	EGSTGHDGA**p**GR	Gadiformes (taxid 8043), hit with collagen 1α1-like of *Gadus morhua* (XP_030195580.1 (same as translated sequence XM_030339720.1)).
Cod (IV)	GPAGAQGGLGA**p**GPK	Gadiformes (taxid 8043), hit with collagen 1α1-like of *Gadus morhua* (XP_030195580.1).
Saithe (IV)	GPAGAQGGLGA**p**GPK	See cod.
Mackerel (III)	GGAGP**p**GATGF**p**GPAGR	Perciformes (taxid 8111), hits with collagen 1α1b of *Sebastes umbrosus*, *Epinephelus lanceolatus* and *Cyclopterus lumpus*.
Rose fish (II)	GG**p**GSSGIAGA**p**GF**p**GSR	Scorpaeniformes (taxid 8111), hit with collagen 1α1b of *Sebastes umbrosus*.
Haddock (IV)	GVTGS**p**GS**p**GPDGK	Gadiformes (taxid 8043), hit with collagen 1α1-like of *Gadus morhua* (XP_030195580.1).
Salmon (II)	GSTGAAGISGA**p**GF**p**GTR	See trout.
Sea bass (II)	GNNGDHGA**p**GPK	Perciformes (taxid 8111), no hits. Hits with collagen 1α3 of *Dicentrarchus labrax* and collagen 1α1b of *Morone saxatilis* when searched against Moronidae (taxid 42148).
Pangasius (II)	GS**p**GPAGITGA**p**GF**p**GTR	Siluriformes (taxid 7995), hit with collagen 1α1b of *Pangasianodon hypophthalmus*.
Tilapia (I)	GA**p**GAAGIAGA**p**GF**p**GAR	Cichliformes (taxid 1489911), hits with collagen 1α1b of *Neolamprologus brichardi*, *Oreochromis aureus*, *Oreochromis niloticus* and with collagen 1α1-like of *Archocentrus centrarchus*.

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	Perciformes	Latidae	33.0	*Toxotes jaculatrix*
Sea bream	Perciformes	Sparidae	47.8	*Acanthopagrus latus*
Trout	Salmoniformes	Salmonidae	59.5	*Oncorhynchus mykiss*
Sardine	Clupeiformes	Clupeidae	11.2	*Carassius auratus*
Hake	Gadiformes	Gadidae	14.6	*Amphiprion ocellaris*
Cod	Gadiformes	Gadidae	15.9	*Epinephelus lanceolatus*
Saithe	Gadiformes	Gadidae	15.6	*Nothobranchius kuhntae* 2
Mackerel	Perciformes	Scombridae	17.7	*Micropterus salmoides*
Rose fish	Scorpaeniformes	Sebastidae	47.5	*Sebastes umbrosus*
Haddock	Gadiformes	Gadidae	12.7	*Nothobranchius kuhntae* 2
Salmon	Salmoniformes	Salmonidae	42.2	*Oncorhynchus mykiss*
Sea bass	Perciformes	Moronidae	48.2	*Morone saxatilis*
Pangasius (gelatin)	Siluriformes	Pangasiidae	62.8	*Pangasianodon hypophthalmus*
Tilapia (gelatin)	Cichliformes	Cichlidae	61.2	*Oreochromis niloticus*
Cod (gelatin)	Gadiformes	Gadidae	13.8	*Nothobranchius kuhntae* 2
Salmon (gelatin)	Salmoniformes	Salmonidae	36.7	*Salvelinus namaycush*

the accession title were considered, it was decided to explore the use 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_030339720.1 (mRNA sequence) and XP_030196341.1 (only protein). Clustal Omega analysis indicated that XM_030339720.1 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_030339720.1) occurring in the cod gelatin sample. The identified peptide GPAGAQGGLGA**p**GPK differs from the analogous part in XP_030196341.1 by 4 amino acids. The elucidated cod peptide sequence only provided a hit with (translated) XM_030339720.1 when searched against taxid 8043 (Gadiformes) using protein blast. Therefore, it is probable that collagen 1α3 was identified in cod. The peptide GPAGAQGGLGA**p**GPK was also detected in a deamidated form, at residue Q6. Accession XM_030339720.1 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 1α3 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 $y_{(n-2)}$/b and $y_{(n-2)}$/a fragment ion series, especially abundant when the third residue is a hydroxyproline (Hyp-3 peptides). Collagen 1α3 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 1α3 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 1α3 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.
Boot-Handford, R. P., Tuckwell, D. S., Plumb, D. A., Rock, C. F., & Poulsom, R. (2003). A novel and highly conserved collagen (proα1(XXVII)) 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.
Breci, L. A., Tabb, D. L., Yates, J. R., & Wysocki, V. K. (2003). Cleavage N-terminal to proline: Analysis of a database of peptide tandem mass spectra. *Analytical Chemistry, 75*, 1963–1971.
Carlson, P., Smalley, D. M., & Van Beneden, R. J. (2013). Proteomic analysis of arsenic-exposed zebrafish (Danio rerio) identifies altered expression in proteins involved in fibrosis and lipid uptake in a gender-specific manner. *Toxicological Sciences, 134*, 83–91.
Duconseille, A., Astruc, T., Quintana, N., Meersman, 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, C., Garrone, R., & Lethias, C. (2002). Evolution of collagens. *The Anatomical Record, 268*, 302–316.
Exposito, J. Y., Larroux, C., Cluzel, C., Valcourt, U., Lethias, 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., & Lethias, C. (2010). The fibrillar collagen family. *International Journal of Molecular Sciences, 11*, 407–426.
Fratzl, P. (2008). *Collagen: 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 Academic Press, Amsterdam. ISBN: 978-0-12-809847-9.
Keen, A. N., Fenna, A. J., McConnell, J. C., Sherratt, M. J., Gardner, P., & Shiels, H. A. (2018). Macro- and micromechanical remodelling in the fish atrium is associated with regulation of collagen 1 alpha 3 chain expression. *Pflügers Archiv, 470*, 1205–1219.
Kimura, S. (1985). The interstitial collagens of fish. *Biology of invertebrate and lower vertebrate collagens*. Boston: Springer.
Kimura, S., Ohno, Y., Miyauchi, Y., & Uchida, N. (1987). Fish skin type I collagen: Wide distribution of an alpha 3 subunit in teleosts. *Comparative Biochemistry and Physiology B, 88*, 27–34.
Kimura, S., & Ohno, Y. (1987). Fish type I collagen: Tissue-specific existence of two molecular forms, $(\alpha1)_2\alpha2$ and $\alpha1\alpha2\alpha3$, in Alaska pollack. *Comparative Biochemistry and Physiology B, 88*, 409–413.
King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., … Rokhsar, D. (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature, 451*, 783–788.
Kleinnijenhuis, A. J., Van Holthoon, F. L., & Herregods, 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*, 106–130.
Kleinnijenhuis, A. J., Van Holthoon, F. L., Maathuis, A. J. H., Vanhoecke, B., Prawitt, J., Wauquier, F., & Wittrant, Y. (2020). Non-targeted and targeted analysis of collagen hydrolysates during the course of digestion and absorption. *Analytical and Bioanalytical Chemistry, 412*, 973–982.
Koli, J. M., Basu, S., Gudipati, V., Chouksey, 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., Killen, A. C., Stockdale, W. T., Potts, H. G., Jefferson, A., … Mommersteeg, M. T. M. (2020). Runx1 promotes scar deposition and inhibits myocardial proliferation and survival during zebrafish heart regeneration. *Development, 147*, dev186569.
Leggatt, R. A., & Iwama, G. K. (2003). Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries, 13*, 237–246.
Makareeva, E., & Leikin, 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., & Bonnaud, L. (2003). Phylogenetic analysis of vertebrate fibrillar collagen locates the position of zebrafish α3(I) 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 chromatographically different α chains. *Biochemistry, 4*, 2590–2596.
Runnegar, 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. ISBN: 978-1-84569-013-7.
Shoulders, M. D., & Raines, R. T. (2009). Collagen structure and stability. *Annual Review of Biochemistry, 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.
Technau, 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., Olijve, J., Brouns, E., & Babel, W. (2016). Gelatin. Ullman's Encyclopedia of Industrial Chemistry. *Wiley-VCH*, 1–22.
Yang, H., Butler, E. R., Monier, S. A., Teubl, J., Fenyö, D., Ueberheide, B., & Siegel, D. (2021). A predictive model for vertebrate bone identification from collagen using proteomic mass spectrometry. *Scientific Reports, 11*, 10900.
Yasui, 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-Handford, R. P., Huxley-Jones, J., Forse, L. N., Mould, A. P., Robertson, D. L., … Sarras, M. P., Jr. (2007). The collagens of hydra provide insight into the evolution of metazoan extracellular matrices. *Journal of Biological Chemistry, 282*, 6792–6802.