

Article

Functional and Structural Properties of Type V Collagen from the Skin of the Shortbill Spearfish (*Tetrapturus angustirostris*)

Qiuyu Han ¹, Tomoyuki Koyama ¹, Shugo Watabe ² and Shoichiro Ishizaki ^{1,*}

¹ Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo 108-8477, Japan; fancy_hqy@163.com (Q.H.)

² School of Marine Biosciences, Kitasato University, Minami, Sagami-hara 252-0373, Kanagawa, Japan

* Correspondence: ishizak@kaiyodai.ac.jp; Tel.: +81-3-5463-0614

Abstract: Type V collagen is considered to be a crucial minor collagen in fish skin with unique physiological functions. In this research, the cDNAs of three procollagens (Tacol5a1, Tacol5a2, and Tacol5a3) in type V collagen were cloned from the skin of shortbill spearfish (*Tetrapturus angustirostris*). The open reading frames (ORFs) of Tacol5a1, Tacol5a2, and Tacol5a3 contained 5991, 4485, and 5607 bps, respectively, encoding 1997, 1495, and 1869 amino acid residues. Each of the deduced amino acid sequences of procollagens contained a signal peptide and a fibrillar collagen C-terminal domain (COLFI). A conserved thrombospondin-like N-terminal domain (TSPN) was found at the N-terminus of Tacol5a1 and 5a3 procollagens, whereas a von Willebrand factor (VWC) was found at the N-terminus of Tacol5a2 procollagen. Tacol5a1, Tacol5a2, and Tacol5a3 had their theoretical isoelectric points of 5.06, 6.75, and 5.76, respectively, and predicted molecular weights of 198,435.60, 145,058.48, and 189,171.18, respectively. The phylogenetic tree analysis revealed that Tacol5a1 of shortbill spearfish clustered with that of yellow perch (*Perca flavescens*) instead of broadbill swordfish (*Xiphias gladius*). In addition, type V collagen was extracted from the shortbill spearfish skin. The *in silico* method demonstrated that shortbill spearfish type V collagen has a high potential for angiotensin-converting enzyme (ACE) inhibition activity (79.50%), dipeptidyl peptidase IV inhibition (74.91%) activity, and antithrombotic activity (46.83%). The structural clarification and possible functional investigation in this study provide the foundation for the applications of exogenous type V collagen derived from fish sources.



Citation: Han, Q.; Koyama, T.; Watabe, S.; Ishizaki, S. Functional and Structural Properties of Type V Collagen from the Skin of the Shortbill Spearfish (*Tetrapturus angustirostris*). *Molecules* **2024**, *29*, 2518. <https://doi.org/10.3390/molecules29112518>

Received: 30 April 2024

Revised: 12 May 2024

Accepted: 22 May 2024

Published: 27 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: type V collagen; cDNA cloning; shortbill spearfish; AlphaFold2; bioactivity

1. Introduction

Collagen, an integral fibrous protein, is ubiquitously found across vertebrate organs, notably enhancing skin and bone tissue functionality [1]. As of now, researchers have identified 29 distinct collagen types [2]. Differentiated into fibrillar and non-fibrillar categories based on their structural and functional roles. Type II, III, IV, and IX collagen have been extensively studied in osteoarthropathies [3,4], Ehlers–Danlos syndrome [5], cardiomyopathic fibrosis [6], ocular diseases [7], Alport syndrome [8,9], and multiple epiphyseal dysplasia [10], among other pathological disorders. Type V collagen, a minor fibrillar collagen subtype isolated through pepsin hydrolysis, plays a critical role in the regulation of procollagen fiber formation within connective tissues expressing type I collagen [11,12]. Comprising three subunits or α chains with similar amino acid sequences, fibrillar collagen is pivotal in tissue architecture and integrity. Initially secreted by cells like fibroblasts and chondrocytes, procollagen undergoes extracellular modifications, including covalent cross-linking and cleavage of peptide bonds, facilitating the transition to mature collagen fibers [13]. In fibrillar collagens, procollagen includes N-terminal and C-terminal prepeptides, N-terminal and C-terminal telopeptides, and a triple-helical structural domain containing the Gly-X-Y repeater. Typically, the X position is proline, and the Y is hydroxyproline.

In fibrous tissues, the composition and ratio of procollagen chains within type V collagen significantly influence health, pathological states, and fibrotic conditions. Markedly increased levels of type V collagen have been identified in tissues compromised by conditions such as cancer, granulation tissue, atherosclerosis, and fibrosis [14]. Mouse embryos lacking the pre- $\alpha 1(V)$ chain show early fatality, highlighting the chain's essential role in development [15]. Notably, mutations in the Col5a1 or Col5a2 genes, which encode the $\alpha 1(V)$ and $\alpha 2(V)$ procollagen chains, respectively, were observed in half of the individuals diagnosed with classical Ehlers–Danlos syndrome (EDS) [16]. Furthermore, elevated expression of the Col5a3 gene, responsible for the $\alpha 3(V)$ procollagen chain, has been inversely associated with gastric cancer survival rates [17]. Nevertheless, mice genetically modified to lack the $\alpha 3(V)$ gene did not show pronounced EDS-like symptoms, but displayed characteristics associated with obesity, including reduced subcutaneous skin layers. This suggests that the $\alpha 3(V)$ chain interacts directly with cellular surface components, influencing islet cell functionality and differentiation, potentially leading to insulin resistance [18].

The shortbill spearfish (*Tetrapturus angustirostris*), a member of the Pacific billfish family, is recognized as one of the largest bony fishes in oceans [19]. As highly migratory apex predators exhibiting solitary behaviors, these fish present substantial research challenges [20]. The harvest of Pacific billfish, including shortbill spearfish, often occurs as a secondary catch in fisheries targeting more valuable species. Despite an overall increase in Pacific billfish catches since 1990, the annual catch rates for shortbill spearfish have consistently declined [21]. Unlike other billfish species such as broad swordfish (*Xiphias gladius*) [19,22], current research on shortbill spearfish is limited.

Marine organisms are esteemed as superior sources of collagen, attributed to their minimal risk of transmitting zoonotic diseases, lack of religious dietary constraints, and significant collagen yields. In the dermal layers of fish, type I and V collagens predominate the fibrillar collagen. Despite the prevalence of type I collagen, the architectural and functional intricacies of type V collagen and its procollagen derivatives from marine origins have remained largely unexplored. This gap underlines the necessity of a thorough investigation into the structural and functional attributes of type V collagen derived from marine species, particularly large fish, to bolster scientific inquiry and application. The purpose of this study is to identify and characterize the structure of type V collagen in fish skin of shortbill spearfish, providing a basis for investigating the potential of type V collagen as a precursor for bioactive peptides.

2. Results and Discussion

2.1. Collagen Identification

Figure 1 displays the SDS-PAGE pattern of type V collagen from the skin of shortbill spearfish. Type V collagen consisted of $\alpha 1(V)$, $\alpha 3(V)$, and $\alpha 2(V)$ chains, aligning with its heterotrimeric structure [23]. In contrast, type I collagen included $\alpha 2(I)$, along with $2\alpha 1(I)$ chains exhibiting twice the intensity of the $\alpha 2(I)$ chain. The $\alpha 3$ chain in type I collagen could not be verified due to the overlapping electrophoretic migration position with those of $\alpha 3(I)$ and $\alpha 1(I)$ [24]. Furthermore, type I and V collagens isolated from shortbill spearfish demonstrated remarkable purity and structural integrity. This was prominently indicated by the absence of any other α subunits in the electrophoretic analyses. Under typical conditions, β and γ bands are visible in the SDS-PAGE patterns of collagens, suggesting the presence of cross-links between α chains. However, such bands were notably absent in the results obtained from these samples. The variability in the degree of collagen cross-linking observed among the samples may be attributable to seasonal variations in fishing practices [25,26]. Specifically, it has been observed that collagens extracted from fish that have endured periods of starvation show a higher degree of cross-linking compared to those derived from well-nourished counterparts, likely due to adaptive biological responses to nutritional stress [27].

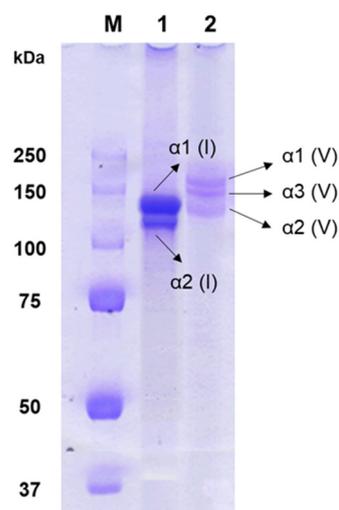


Figure 1. SDS-PAGE patterns of type I and V collagens isolated from shortbill spearfish. Lanes M, 1, and 2 indicate molecular weight markers, type I collagen, and type V collagen, respectively.

Type V collagen consisted of $\alpha 1(V)$, $\alpha 3(V)$, and $\alpha 2(V)$ chains with molecular weights of 154 kDa, 145 kDa, and 127 kDa, respectively, which were larger than those of corroborating findings of type I collagens with $\alpha 1(I)$ and $\alpha 2(I)$ chains with molecular weights of 130 kDa and 120 kDa, respectively, from prior studies on shortbill spearfish skin [28,29]. These results suggest that type V collagen has extensive intra- and/or intermolecular cross-linking, more than that of type I collagen. The SDS-PAGE pattern distinctly revealed three unique procollagens within type V collagen, laying the groundwork for subsequent investigations.

2.2. Physicochemical Properties of Shortbill Spearfish Type V Procollagens

Degenerate primers were constructed for the PCR reaction, which was conserved at the highest similarity among the nearest genetic relationship of bony fish $\alpha 1(V)$, $\alpha 2(V)$, and $\alpha 3(V)$ procollagens. The sequence of obtained cDNA fragments was verified using the BLAST in the NCBI database. The cDNAs encoding Tacol5a1, Tacol5a2, and Tacol5a3 procollagens were cloned successfully. The sequence analysis showed that ORFs of cDNAs encoding Tacol5a1, Tacol5a2, and Tacol5a3 consisted of 5991, 4485, and 5607 bps, respectively, encoding 1997, 1495, and 1869 amino acid residues with a theoretical isoelectric point of 5.06, 6.75, and 5.76, respectively (Table 1). The calculated instability indexes of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens were 32.61, 26.47, and 30.56, respectively, indicating that the procollagens are stable (Table 1). The determined nucleotide sequences of cDNAs encoding Tacol5a1, Tacol5a2, and Tacol5a3 procollagens have been deposited in the NCBI database under the accession numbers OR700193, OR700194, and OR700195, respectively.

Table 1. Physicochemical properties of shortbill spearfish Tacol5a1, Tacol5a2, and Tacol5a3 procollagens.

Encoding Gene Name	Amino Acid Residues	Molecular Mass (kDa)	pI	Instability Index
Tacol5a1	1997	198.44	5.06	32.61
Tacol5a2	1495	145.06	6.75	26.47
Tacol5a3	1869	189.15	5.76	30.56

The heatmaps of amino acid compositions of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens are shown in Figure 2. In the compositions, glycine (Gly) was the most abundant (22–27%), followed by proline (Pro) (12–17%), indicating a rich GC content in the amino acid sequence of procollagens [30]. This typical structure of collagen chains determines the regular structure and contributes to the formation of stabilizing bonds [31]. Meanwhile, cysteine (Cys) and tryptophan (Trp), which are not supposed to be in the triple-helix

structure [24], were present at less than 1%. The results are similar to the amino acid compositions of α -chains in type I and V collagens isolated from other species [31–34].

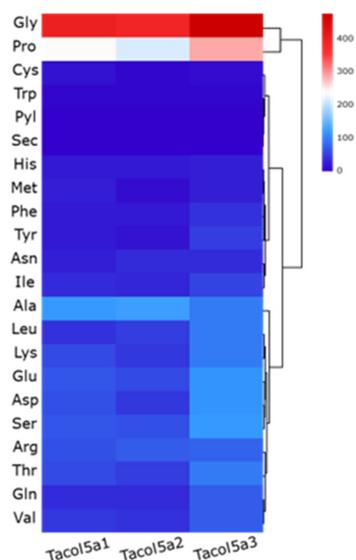


Figure 2. Heatmaps of deduced amino acids compositions of shortbill spearfish type V procollagens of Tacol5a1, Tacol5a2, and Tacol5a3.

The contents of Gly-Pro-Pro (GPP) and Gly-Gly (GG) motifs in the triple-helical domains of shortbill spearfish Tacol5a1, Tacol5a2, and Tacol5a3 procollagens are shown in Table 2. The GPP motif contents in the triple-helical domains of shortbill spearfish type V procollagens of Tacol5a1, Tacol5a2, and Tacol5a3 (43, 35, and 31, respectively) were higher than those of shortbill spearfish type I procollagens of Tacol1a1 and Tacol1a2 (29 and 26, respectively). With the exception of the Tacol5a2 procollagen, aquatic animals contained the levels of the GPP motif lower than those of terrestrial animals. The content of the GPP motif was similar among animals from similar living environments. Total GPP motif content is the major factor influencing the thermal stability of collagen [35]. Estimating the Gly-Pro-Hyp content in mature proteins is feasible through analysis of the GPP content within the cDNA sequence, which could be used to predict the thermal stability of collagen [36,37]. Therefore, the higher GPP contents of shortbill spearfish type V procollagen indicated a more stable structure of type V collagen than type I collagen. This result is consistent with the thermal behavior reported by Wang et al. [29]. The Tacol5a1 procollagen showed the highest stability, which was associated with a high cross-linking of $\alpha 1(V)$ [38]. However, the lower GPP content of shortbill spearfish type V procollagen compared with those of terrestrial animals indicates a lower denaturation temperature, which may be due to their cold-water habitat. On the other hand, the GG motif could be responsible for the partial skewing of the triple-helix structure and reduced thermal stability [32,37]. The lower GG motif content is also attributed to the higher thermal stability of shortbill spearfish type V collagen compared to zebrafish and sailfish.

Table 2. Gly-Pro-Pro and Gly-Gly contents in the triple helical domain of shortbill spearfish Tacol5a1, Tacol5a2, and Tacol5a3 procollagens in comparison with those of other animals.

	Triple Helix Region Length	Gly-Pro-Pro Content	Gly-Gly Content	Accession Number
$\alpha 1(I)$ and $\alpha 2(I)$ procollagens				
Shortbill spearfish Tacol1a1	1014	29	13	OR700191
Shortbill spearfish Tacol1a2	1015	26	14	OR700192
$\alpha 1(V)$ procollagen				
Shortbill spearfish Tacol5a1	1014	43	4	
Zebrafish	1014	42	3	ADG36303.1
Broadbill Swordfish	1014	43	6	XP040014133.1
Human	1014	49	7	NP000084.3
Mouse	1014	44	7	EDL08374.1
Chicken	1014	48	3	NP990121.2
Shortbill spearfish $\alpha 2(V)$ procollagen				
Tacol5a2	1017	35	6	
Zebrafish	1017	34	11	NP001139254.1
Broadbill Swordfish	1017	32	4	XP040004036.1
Human	1017	30	6	NP000384.2
Mouse	1017	30	6	NP031763.2
Chicken	1017	30	11	XP040532372.1
$\alpha 3(V)$ procollagen				
Shortbill spearfish Tacol5a3	1011	31	11	
Zebrafish	1011	32	11	NP001177685.1
Broadbill Swordfish	1011	30	10	XP_040013352.1
Human	1011	41	4	NP001845.3
Mouse	1011	43	5	AAF59901.1
Chicken	1011	40	6	XP422303.4

2.3. Primary Structure Analysis of Shortbill Spearfish Type V Procollagens

Regulation of mRNA transcription is critically influenced by the triple-helix structure [39]. The deduced amino acid sequences of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens from shortbill spearfish are shown in Figure 3. Tacol5a2 procollagen contained one internal coupling site (GMKGHR). These interleaved coupling structures help stabilize the collagen fiber structure [40]. Furthermore, Tacol5a1, Tacol5a2, and Tacol5a3 procollagen contained 1, 6, and 4 arginine–glycine–aspartate (RGD) cell adhesion sites, respectively. It is noteworthy that potential Asn-X-Thr/Ser glycosylation sites are present in all of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens (indicated by boxes). Among them, the glycosylation site of Tacol5a2 procollagen was found only in the C-terminal non-triple-helical structure. Tacol5a2 procollagen features an amidation reaction site (EGKR), a characteristic not widely conserved across different fish species. Consequently, while the amino acid sequence of shortbill spearfish type V procollagens shared similarities with those of other fish, notable distinctions were also present. Moreover, the putative N- and C-proteinase

cleavage sites were shown, which were based on data from the other vertebrate procollagen chains [41]. In the N-terminus of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens, the signal peptides with 33 amino acid residues (¹MDTHIRWVKVRRIRDVQITLAVVLLFVISQASS³³), 25 amino acid residues (¹MMS-FVHLRTRFLFLVVSVAQVLIVTC²⁵), and 28 amino acid residues (¹MDHLIRTRSRRIPLFLLI-LLHVTTTQA²⁸) were observed, respectively.

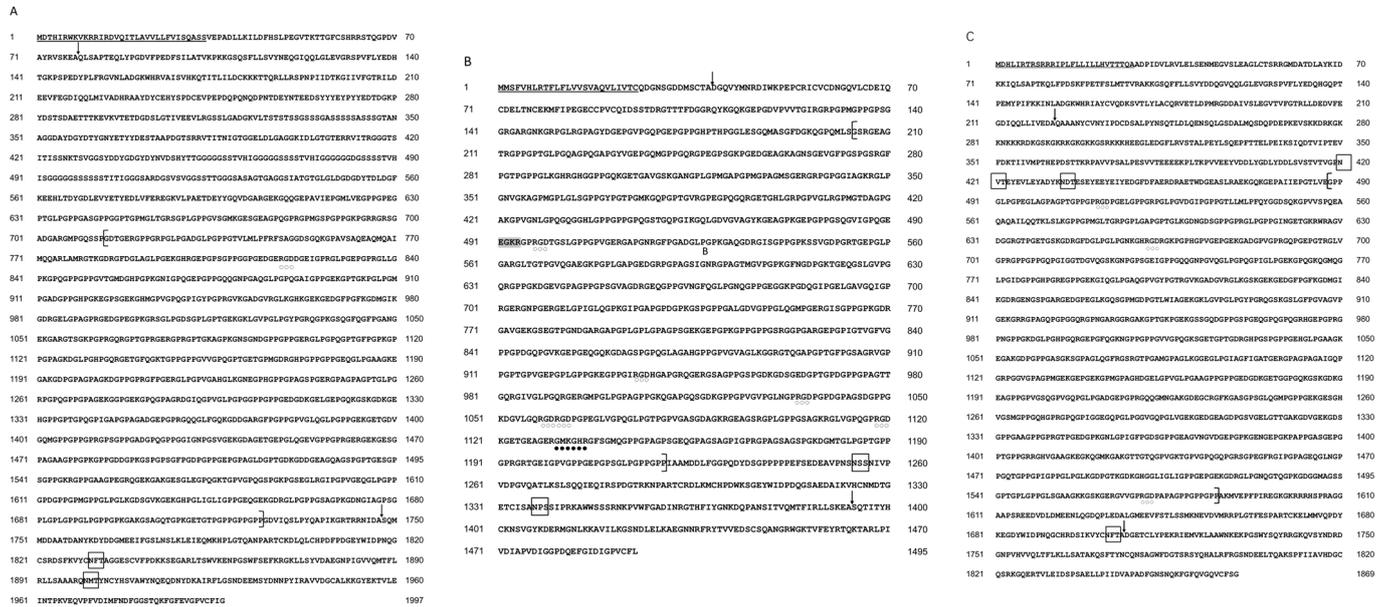


Figure 3. The deduced amino acid sequences of shortbill spearfish Tacol5a1 (A), Tacol5a2 (B), and Tacol5a3 procollagens (C). A single underlined letter indicates the cleavage site of the putative signal peptide, arrows indicate cleavage sites of N-propeptides and C-propeptides, and the putative intermolecular cross-linking sites are marked by black closed circles. The RGD sites that represent the potential cell adhesion sites are marked by open circles, Asn-X-Thr/Ser indicates glycosylation sites are boxed, and a shadow indicates the amidation site.

According to the identification of conserved domains, each of the deduced amino acid sequences of procollagens contained a conservative domain of the fibrillar collagen C-terminus (COLFI), suggesting that shortbill spearfish type V collagens produced from Tacol5a1, Tacol5a2, and Tacol5a3 procollagens are members of the fibrillar collagen family (Figure 4). Moreover, the thrombospondin-like N-terminal domains (TSPN) related to the heparin-binding and cell adhesion domain of thrombospondin were found in Tacol5a1 and Tacol5a3 procollagens. The presence of TSPN suggested the potential of Tacol5a1 and Tacol5a3 procollagens to influence the physiology and pathology of cardiovascular disease [13]. Furthermore, the von Willebrand factor (VWC) domains [42] were observed at the N-terminus of Tacol5a2 procollagen, potentially suggesting an impact on platelet function.

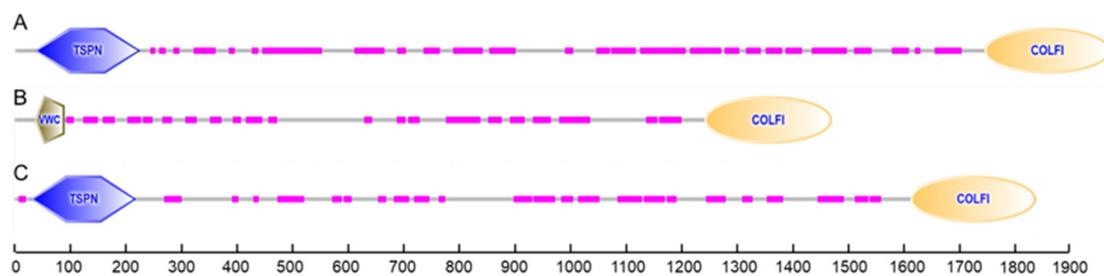


Figure 4. The prediction of conserved domains of shortbill spearfish Tacol5a1 (A), Tacol5a2 (B), and Tacol5a3 (C) procollagens.

2.4. Secondary and Tertiary Structure Prediction of Shortbill Spearfish Type V Collagens

As shown in Figure 5 and Table 3, the contents of α -helix, β -sheet, and turn are concentrated at the N- and C-termini of shortbill spearfish type V procollagens. Tacol5a1 procollagen showed the highest beta sheet content (13.32%), and lower coil content (71.06%), indicating the higher stability. These results are consistent with the previous discussion of the GPP motif content, which could account for the stable structural stability. Furthermore, alpha helix and beta sheet were mainly distributed in the C- and N-termini of procollagens, and most of the triple-helix domains were coiled, which is also a special characterization of collagen structure.

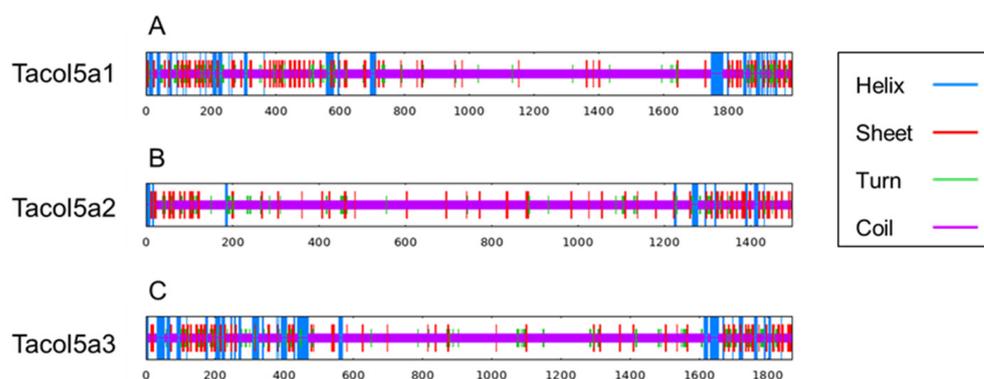


Figure 5. The prediction of the secondary structure of shortbill spearfish Tacol5a1 (A), Tacol5a2 (B), and Tacol5a3 (C) procollagens. The blue line represents the α -helix, the red line represents the extended strand, and the purple line represents the random coil.

Table 3. Prediction of secondary structure composition of shortbill spearfish Tacol5a1, Tacol5a2, and Tacol5a3 procollagens.

%	Alpha Helix	Beta Sheet	Turn	Coil
Tacol5a1	9.81	13.32	5.81	71.06
Tacol5a2	3.81	11.64	4.75	79.80
Tacol5a3	12.52	11.56	6.37	69.56

The tertiary structure prediction of shortbill spearfish procollagens by AlphaFold 2 based on Colab is shown in Figure 6, suggesting that the procollagens are spheroidal (Figure 6a–c). Hence, the Colab-based AlphaFold2 could not predict the 3D structure very accurately for collagen. Figure 6d,e show the C- and N-terminal structural domain comparisons, respectively. Quantitative assessment of similarity between two protein structures was achieved by calculating RMSD after superimposing them in PyMOL software. In Figure 6d, the RMSD values between the two procollagens were small (<1.35), indicating a high level of structural similarity [43]. Further, since no conserved domains were found at the N-terminus of Tacol5a2 procollagen in Figure 4, only the N-terminal region without the triple-helix domain of Tacol5a1 and Tacol5a3 procollagens are shown in Figure 6e. The RMSD level of the structures of Tacol5a3 and Tacol5a1 procollagens was 4.797, indicating a structural similarity of less than 50% [44]. However, the removal of the N-terminal telopeptide reduced the RMSD level to 0.560, indicating the structure similarity is higher than 80% (Figure 6f). These results suggest that Colab-based AlphaFold2 still needs further deep learning on the tertiary organization prediction of proteins with high coil content and a special structure.

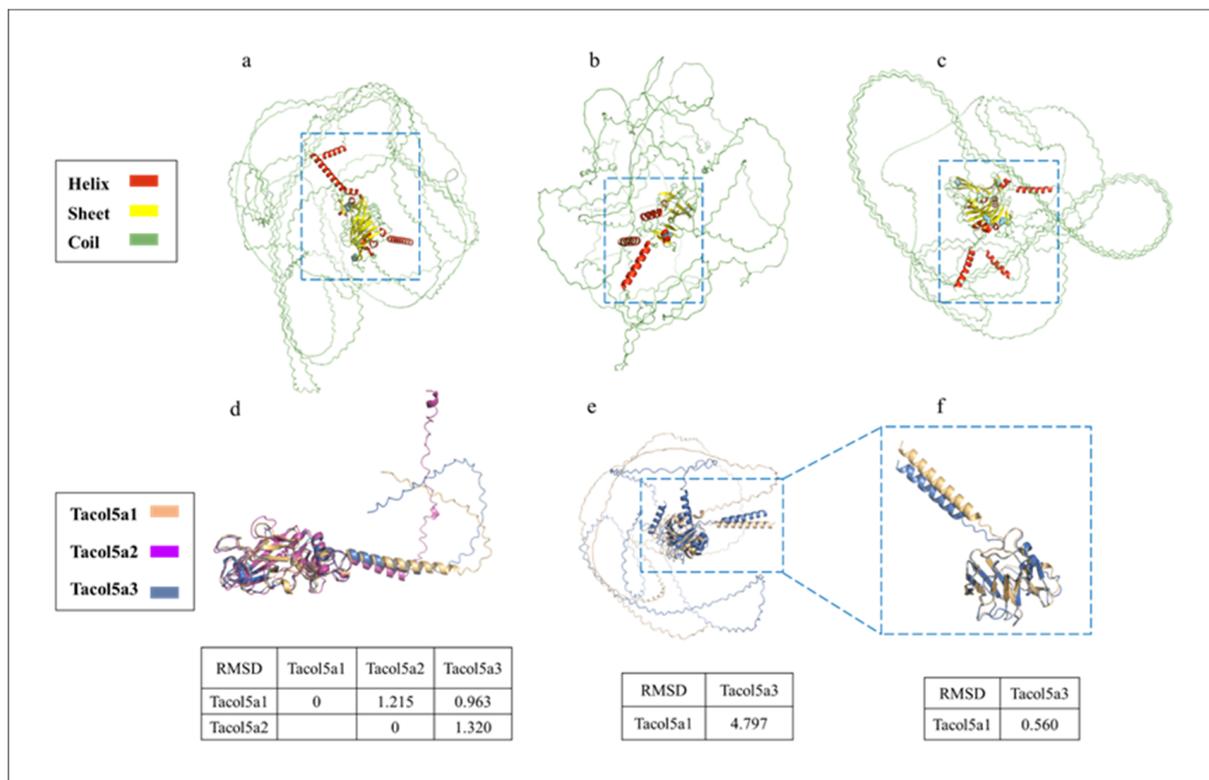


Figure 6. Tertiary structure prediction of shortbill spearfish Tacol5a1, Tacol5a2, and Tacol5a3 procollagens. Tertiary structure prediction of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens are shown in panels (a–c); the C-terminus and N-terminus without triple-helix structures are boxed; (d) indicates an alignment of the tertiary structure of the C-terminus; (e) denotes an alignment of the tertiary structure of Tacol5a1 and Tacol5a2 procollagens without triple-helix structural domains; (f) denotes an alignment of the structure of the N-terminus without telopeptide.

2.5. Multiple Sequence Alignment and Phylogenetic Analysis

Figure 7 presents the comparative homology of amino acid sequences for procollagens across various species, revealing significant conservation between shortbill spearfish procollagens and those from other species. Figure 7A–C illustrate that the degree of homology for type V procollagens varies across species, with Tacol5a1 procollagen demonstrating the most substantial sequence conservation. Notably, the identities between all three shortbill spearfish procollagens and that of broadbill swordfish exceed 90%. Specifically, the identities are approximately 89% for Tacol5a1 procollagen, 74% to 83% for Tacol5a2 procollagen, and 67% to 84% for Tacol5a3 procollagen. Remarkably, Tacol5a1, Tacol5a2, and Tacol5a3 procollagens shared a high degree of identities with broadbill swordfish (97.14%, 90.56%, and 93.97%, respectively).

Additionally, the phylogenetic tree of shortbill spearfish is shown in Figure 7D, which was constructed via a multi-locus sequence analysis. All examined sequences were clustered into Col5a1, Col5a1, or Col5a3, with a topology that aligned with the established phylogeny of bony fish. Notably, the closest homology observed was between shortbill spearfish and broadbill swordfish, highlighting a distinct evolutionary lineage among bony fish, amphibians, reptiles, birds, and mammals.

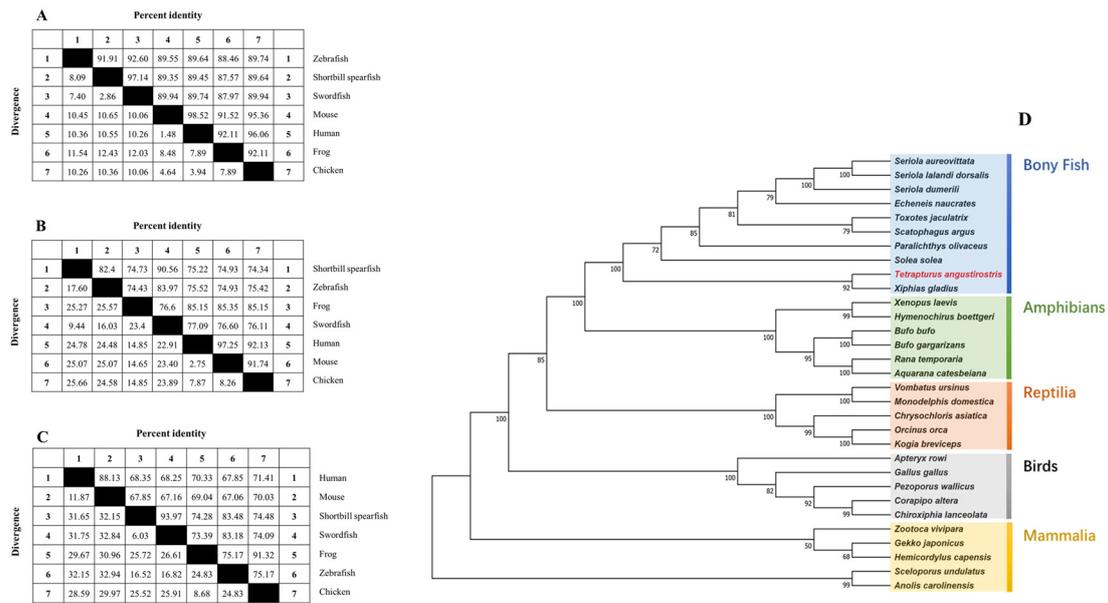


Figure 7. The homology of shortbill spearfish Tacol5a1 (A), Tacol5a2 (B), and Tacol5a3 (C) procollagens with other procollagens and their phylogenetic relationship (D). (D) is a maximum likelihood (ML) tree based on the amino acid sequences of procollagens. The other amino acid sequences were downloaded from the NCBI protein database. The phylogenetic tree was constructed by joining amino acid sequences of various species. Numbers at nodes indicate bootstrap values for 1000 replicates.

2.6. Potential Bioactivity of Type V Collagen by In Silico Method

All the bioactivities in the database were selected to evaluate the feasibility of the three shortbill spearfish type V procollagens all together as a precursor for bioactive peptides by comparing the A values [45]. The A value denotes the frequency of occurrence of the bioactive fragments in collagen. A higher A value represents the higher potential of the protein to produce bioactive fragments. The frequency of occurrence for bioactive fragments in each bioactive type is shown in Figure 8. It is evident from this figure that fragments with dipeptidyl peptidase IV inhibition activity (74.91%), ACE inhibitory activity (75.90%), and thrombin inhibitory activity (46.83%) dominate the sequences of all type V procollagens. This suggests a significant presence of these activities within the collagen sequence. The BIOPEP-UWM database contains 48 major types of peptide bioactivities, of which 34 subclasses of peptide bioactivities could be identified in the obtained type V collagen amino acid sequence, indicating that type V collagen is a valuable potential candidate for the production of bioactive peptides.

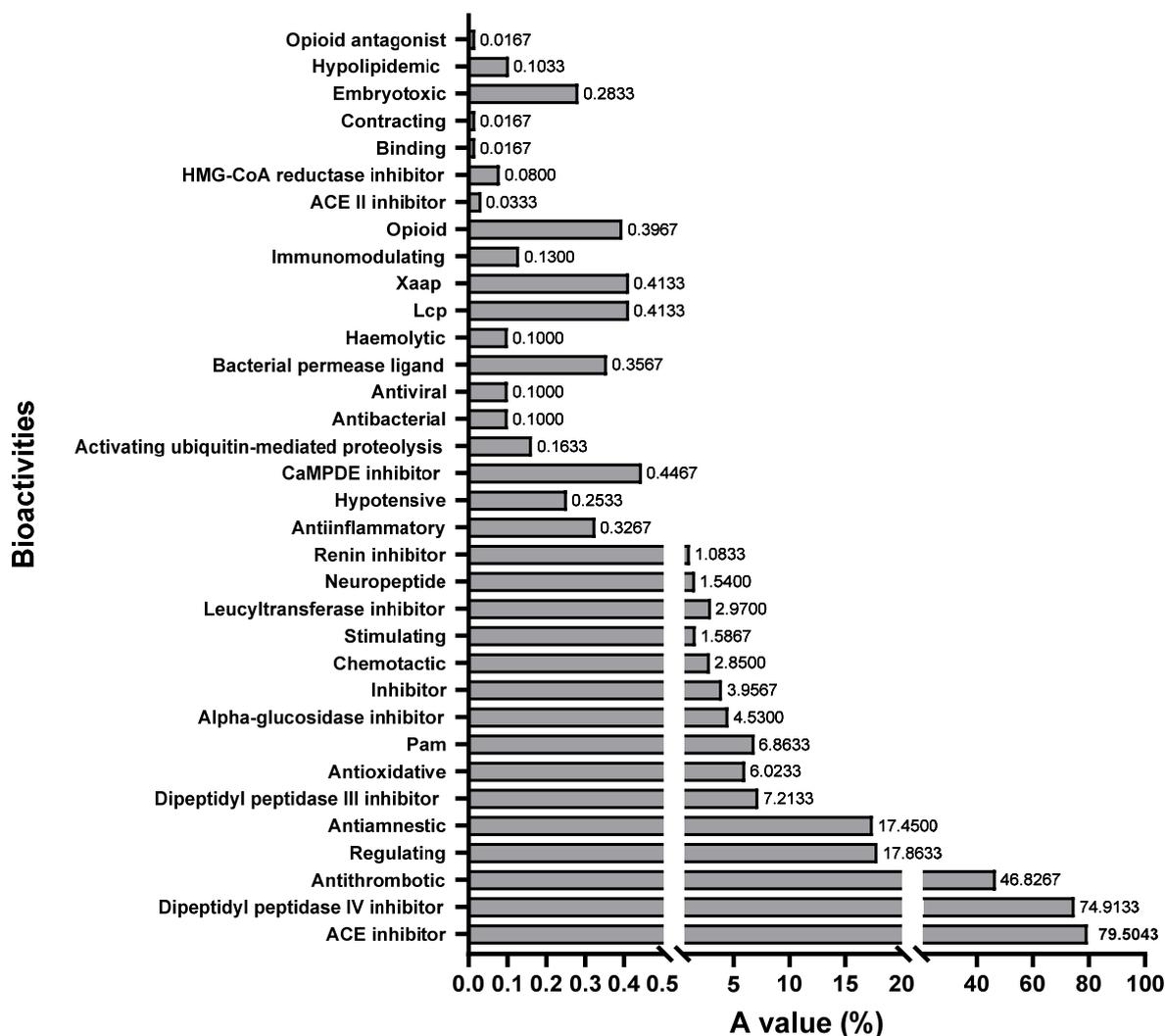


Figure 8. The frequency of occurrence of peptides with a given activity (A value) in type V collagen of shortbill spearfish.

3. Materials and Methods

3.1. Materials

Shortbill spearfish specimens (body weight: 12–18 kg) were purchased from fishermen in Kesen-Numa City, Miyagi Prefecture, Japan. They were frozen immediately after being caught and transported to the laboratory of Tokyo University of Marine Science and Technology. Pepsin (1:10,000) (from Porcine Stomach Mucosa, EC 3.4.23.1), NaOH, acetic acid, HCl, NaCl, tris-(hydroxymethyl)-aminomethane (Tris), butanol, and bromophenol blue were purchased from FUJIFILM Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Ultrapure water was prepared by Milli-Q system (Millipore, Tokyo, Japan).

3.2. Fish Skin Pretreatment

The fish skin was washed after removing the flesh and cut into small pieces (less than $0.5 \times 0.5 \text{ cm}^2$). A solution of 0.1 M of NaOH was used to soak the skin at a ratio of 1:30 (*w/v*) at 4 °C for 48 h in order to remove the non-collagenous protein, refreshing the alkaline medium solution every 6 h. The skin pieces were cleaned to a neutral pH with cold water, followed by treatment with 10% butanol. Concurrently, non-collagenous proteins and alkaline-soluble collagens were eliminated, and the remaining materials were rinsed with cold ultrapure water until it reached a neutral to slightly basic pH level. The defatted skin was cleaned and stored at -25 °C . Before use, the skin was chilled at -85 °C for 6 h

and then minced with a grinder (SKF-H100, Tiger Magic Bottle Co., Ltd., Osaka, Japan). All steps were conducted at 4 °C, with continuous agitation using a magnetic stirrer.

3.3. Preparation of Type V Collagen

The extraction of distinct collagen types was performed using the method described by Han et al. [28]. The pretreated skin was agitated in 0.5 M acetic acid containing 0.1% pepsin (*w/v*) at a 1:20 (*w/v*) ratio for 48 h, then centrifuged at 10,000× *g* for 60 min. The supernatant underwent salting out by the addition of NaCl to achieve a final concentration of 1.2 M, followed by centrifugation at 10,000× *g* for 60 min. The resulting precipitate was redissolved in 0.5 M Tris-HCl buffer (pH 7.5), and the NaCl concentration was subsequently adjusted to 4.0 M. The precipitate was redissolved and then added with 2.4 M NaCl to separate the type V and type I collagens into supernatant and precipitate fractions, respectively. The supernatant was dialyzed against ultrapure water with dialysis membranes (MWCO:12–14 kDa, standard RC tubing, Repligen Corp., Waltham, MA, USA). The precipitate was dissolved into 0.5 M acetic acid and dialyzed against 0.1M acetic acid and ultrapure water, successively. The pepsin-soluble collagen solutions were stored at −25 °C after lyophilization. All procedures were performed at 4 °C.

3.4. SDS–Polyacrylamide Gel Electrophoresis (SDS–PAGE) Pattern

SDS–PAGE of pepsin-soluble collagen extracted from the fish skin was performed according to the method of Laemmli [46] with some modifications. A sample at 2 mg/mL in 0.5 M acetic acid was treated with 2× sample loading buffer (60 mM Tris–HCl, pH 8.0, containing 25% glycerol, 2% SDS, 0.1% bromophenol blue) in the presence of 2% β-mercaptoethanol. SDS–PAGE of pepsin-soluble collagen was conducted on 8% resolving gel and 5% stacking gel on a cPAGE Ace Twin (WSE-1025W, ATTO Co., Ltd., Tokyo, Japan). After electrophoresis, the gel was stained with 0.1% (*w/v*) Coomassie blue R-250 and then destained.

3.5. cDNA Cloning of Procollagens

3.5.1. RNA Extraction and cDNA Synthesis

The skin was dissected out from shortbill spearfish, cut into small pieces, and transformed into liquid nitrogen immediately. Frozen tissues were put into a 50 mL tube and homogenized with TRIzol™ Reagent (Invitrogen, Carlsbad, CA, USA) to isolate total RNA. Poly A⁺ RNA was isolated and purified from total RNA using illustra™ QuickPrep Micro mRNA Purification Kit (GE Healthcare, Tokyo, Japan) according to the manufacturer's instructions. The quality, purity, and integrity of RNA were assessed by the A260/280 ratio.

Synthesis of double-stranded cDNA (ds cDNA) was performed with Marathon® cDNA Amplification Kit (Clontech Lab, Mountain View, CA, USA) according to the manufacturer's introductions.

3.5.2. Cloning of cDNA Encoding Procollagens

The gene-specific primers (GSPs) of procollagens nucleotide sequences (Tacol5a1, Tacol5a2, and Tacol5a3) were constructed based on the conserved regions selected from broadbill swordfish *Xiphias gladius* (XM040158199.1, XM040148102.1, XM040157418.1), mandarin fish (*Siniperca chuatsi*) (XM044175423.1, XM044215955.1, XM044180658.1), redfin perch (*Perca fluviatilis*) (XM039779366.1, XM039817710.1, XM039825796.1), leopard coral grouper (*Plectropomus leopardus*) (XM042508952.1, XM042494419.1, XM042504858.1), yellow-tail amberjack *Seriola (lalandi dorsalis)* (XM023422226.1, XM023423150.1, XM023401417.1), barramundi perch (*Lates calcarifer*) (XM051072972.1, XM018683186.2, XM018705093.2), and banded archerfish (*Toxotes jaculatrix*) (XM041058269.1, XM041056572.1, XM041061690.1) using the ClustalW program (<https://www.genome.jp/tools-bin/clustalw>, accessed on 19 March 2023) (Supplementary Materials).

The cDNA sequences encoding Tacol5a1, Tacol5a2, and Tacol5a3 were obtained using the template diluted 250 times from the adapted ds cDNA. The PCR reaction was carried

out in a total volume of 25 μ L using Ex Taq DNA polymerase (Takara, Otsu, Japan). PCR thermal cycling conditions were 30 cycles of 98 °C for 10 sec for denaturation, 53–55 °C 30 sec for annealing, and 72 °C for 1 sec for extension, with a final extension step at 72 °C for 3 min. The detailed annealing temperature depended on each primer pair. The cDNAs of Tacol5a1, Tacol5a2, and Tacol5a3 were obtained using the sets of adaptor primers (AP1: CCATCCTAATACGACTCACTATAGGGC, AP2: ACTCACTATAGGGCTCGAGCGGC) with GSP-AR (5'-end) or GSP-AF (3'-end) (Supplementary Materials). PCR products were gel-purified by FastGene™ Gel/PCR Extraction Kit (Nippon Gene, Tokyo, Japan), sub-cloned into the pGEM-T easy vector (Promega, Madison, WI, USA), treated with a BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA), and sequenced with an ABI 3130 Genetic Analyzer (Applied Biosystems). The cDNA sequences of Tacol5a1, Tacol5a2, and Tacol5a3 were generated by overlapping the fragments.

3.6. Bioinformatics Analysis

The cDNA sequences and deduced amino acid sequences were analyzed using SnapGene software (www.snapgene.com, accessed on 1 April 2023). The open reading frames (ORFs) were identified using the NCBI ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>, accessed on 1 August 2023). Sequence similarity was explored using the BLAST tool (<http://blast.ncbi.nlm.nih.gov/>, accessed on 1 August 2023). Physical and chemical parameter calculations were performed using the ExPasy-ProtParam tool (<https://web.expasy.org/protparam/>, accessed on 5 August 2023). Secondary structure prediction was carried out using ESPript V3.0 (<https://espript.ibcp.fr/ESPrript/ESPrript/>, accessed on 7 August 2023). The prediction of conserved domains in the amino acid sequence was performed using the SMART program (<http://smart.embl-heidelberg.de/>, accessed on 7 August 2023). The signal peptides were determined using the software SignalP 4.0 (<https://services.healthtech.dtu.dk/services/SignalP-4.1/>, accessed on 8 August 2023). ColabFold V1.4 was used to model the tertiary structure of Tacol5a1, Tacol5a2, and Tacol5a3 procollagens [47]. To quantitatively assess the similarity between two protein structures, the root mean square deviation (RMSD) was calculated by superimposing the structures using PyMOL Molecular Graphics System v2.4.0 (Schrodinger, LLC.; New York, NY, USA) [48]. A phylogenetic neighbor-joining tree was constructed using Mega11 V11.0.13 [49], and 1000 bootstrap trials were conducted to increase the confidence values for the resulting phylogenetic tree.

3.7. Assessment of Bioactive Peptides in Type V Collagen

The assessment of potential bioactive peptides in type V collagen was carried out using BIOPEP-UWM analysis (<https://biochemia.uwm.edu.pl/en/biopep-uwm-2/>, visited on January 2024) [50]. The amino acid sequence of each procollagen chain was subjected to “profiles of potential biological activity”. The potential bioactive peptides from type V collagen derived from shortbill spearfish were screened. The frequency of fragments with bioactivities in protein sequences, denoted as A%, was characterized using the following equation:

$$A(\%) = \frac{a}{N} \times 100$$

where a is the number of fragments and N is the number of amino acid residues of the protein.

4. Conclusions

This study marks the cloning of cDNAs encoding type V procollagen from shortbill spearfish skin. The findings highlight Tacol5a1, Tacol5a2, and Tacol5a3 procollagens for exceptional thermal stability. Analysis of the deduced amino acid sequences revealed that shortbill spearfish type V collagens shared fundamental biological functions with those in other bony fish (more than 90%). Additionally, *in silico* analysis confirmed the potential of shortbill spearfish type V collagens to generate bioactive fragments. The investigation

of the bioactivity of shortbill spearfish type V collagens is the forthcoming work in our research. Improving the accuracy of collagen tertiary structure prediction will also be a future research goal.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules29112518/s1>, Table S1: Primer names and sequence information. Figure S1: Primers location at each coding regions of shortbill spearfish procollagens.

Author Contributions: Conceptualization, Q.H. and S.I.; investigation, Q.H.; writing—original draft preparation, Q.H. and S.I.; writing—review and editing, T.K., S.I., and S.W.; supervision, S.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JST SPRING, grant number JPMJSP2147 (to Q.H.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shoulders, M.D.; Raines, R.T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958. [[CrossRef](#)] [[PubMed](#)]
2. Safandowska, M.; Pietrucha, K. Effect of fish collagen modification on its thermal and rheological properties. *Int. J. Biol. Macromol.* **2013**, *53*, 32–37. [[CrossRef](#)] [[PubMed](#)]
3. Harris, R.B.; Fonseca, F.L.A.; Sharp, M.H.; Ottinger, C.R. Functional characterization of undenatured type II collagen supplements: Are they interchangeable? *J. Diet. Suppl.* **2022**, *19*, 717–732. [[CrossRef](#)] [[PubMed](#)]
4. Xu, R.; Zheng, L.; Su, G.; Luo, D.; Lai, C.; Zhao, M. Protein solubility, secondary structure and microstructure changes in two types of undenatured type II collagen under different gastrointestinal digestion conditions. *Food Chem.* **2021**, *343*, 128555. [[CrossRef](#)] [[PubMed](#)]
5. Liu, J.; Xu, M.; Wu, J.; Zhang, H.; Yang, L.; Lun, D.; Hu, Y.; Liu, B. Picrosirius—Polarization Method for Collagen Fiber Detection in Tendons: A Mini—Review. *Orthop. Surg.* **2021**, *13*, 701–707. [[CrossRef](#)] [[PubMed](#)]
6. Rezvani Ghomi, E.; Nourbakhsh, N.; Akbari Kenari, M.; Zare, M.; Ramakrishna, S. Collagen—based biomaterials for biomedical applications. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **2021**, *109*, 1986–1999. [[CrossRef](#)] [[PubMed](#)]
7. Song, Y.; Overmass, M.; Fan, J.; Hodge, C.; Sutton, G.; Lovicu, F.J.; You, J. Application of collagen I and IV in bioengineering transparent ocular tissues. *Front. Surg.* **2021**, *8*, 639500. [[CrossRef](#)] [[PubMed](#)]
8. Warady, B.A.; Agarwal, R.; Bangalore, S.; Chapman, A.; Levin, A.; Stenvinkel, P.; Toto, R.D.; Chertow, G.M. Alport syndrome classification and management. *Kidney Med.* **2020**, *2*, 639–649. [[CrossRef](#)] [[PubMed](#)]
9. Quinlan, C.; Rheault, M.N. Genetic basis of type IV collagen disorders of the kidney. *Clin. J. Am. Soc. Nephrol.* **2021**, *16*, 1101–1109. [[CrossRef](#)]
10. Dennis, E.P.; Greenhalgh—Maychell, P.L.; Briggs, M.D. Multiple epiphyseal dysplasia and related disorders: Molecular genetics, disease mechanisms, and therapeutic avenues. *Dev. Dyn.* **2021**, *250*, 345–359. [[CrossRef](#)]
11. Burgeson, R.E.; El Adli, F.A.; Kaitila, I.I.; Hollister, D.W. Fetal membrane collagens: Identification of two new collagen alpha chains. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 2579–2583. [[CrossRef](#)] [[PubMed](#)]
12. Birk, D.E.; Brückner, P. Collagens, suprastructures, and collagen fibril assembly. In *The Extracellular Matrix: An Overview*, Springer: Berlin/Heidelberg, Germany, 2010; pp. 77–115.
13. Vogel, W.F. Collagen-receptor signaling in health and disease. *Eur. J. Dermatol.* **2001**, *11*, 506–514. [[PubMed](#)]
14. Mak, K.M.; Png, C.Y.M.; Lee, D.J. Type V collagen in health, disease, and fibrosis. *Anat. Rec.* **2016**, *299*, 613–629. [[CrossRef](#)] [[PubMed](#)]
15. Wenstrup, R.J.; Florer, J.B.; Brunskill, E.W.; Bell, S.M.; Chervoneva, I.; Birk, D.E. Type V collagen controls the initiation of collagen fibril assembly. *J. Biol. Chem.* **2004**, *279*, 53331–53337. [[CrossRef](#)] [[PubMed](#)]
16. Leeming, D.; Karsdal, M. Type V collagen. In *Biochemistry of Collagens, Laminins and Elastin*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 55–60.
17. Tan, Y.; Chen, Q.; Xing, Y.; Zhang, C.; Pan, S.; An, W.; Xu, H. High expression of COL5A2, a member of COL5 family, indicates the poor survival and facilitates cell migration in gastric cancer. *Biosci. Rep.* **2021**, *41*, BSR20204293. [[CrossRef](#)] [[PubMed](#)]
18. Huang, G.; Ge, G.; Wang, D.; Gopalakrishnan, B.; Butz, D.H.; Colman, R.J.; Nagy, A.; Greenspan, D.S. $\alpha 3$ (V) collagen is critical for glucose homeostasis in mice due to effects in pancreatic islets and peripheral tissues. *J. Clin. Invest.* **2011**, *121*, 769–783. [[CrossRef](#)] [[PubMed](#)]

19. Rick, T.; Harvey, V.L.; Buckley, M. Collagen fingerprinting and the chumash billfish fishery, Santa Barbara Channel, California, USA. *Archaeol. Anthropol. Sci.* **2019**, *11*, 6639–6648. [[CrossRef](#)]
20. Pepperell, J.G. *Movements and Variations in Early Year Class Strength of Black Marlin, Makaira Indica Off Eastern Australia*; Fisheries Research Institute: Keelung, Taiwan, 1988.
21. Gentner, B. *Economic Analysis of International Billfish Markets*; Gentner Consulting Group: Silver Spring, MD, USA, 2007.
22. Cobas, N.; Gómez-Limia, L.; Franco, I.; Martínez, S. Amino acid profile and protein quality related to canning and storage of swordfish packed in different filling media. *J. Food Compos. Anal.* **2022**, *107*, 104328. [[CrossRef](#)]
23. Nishimoto, M.; Sakamoto, R.; Mizuta, S.; Yoshinaka, R. Identification and characterization of molecular species of collagen in ordinary muscle and skin of the Japanese flounder *Paralichthys olivaceus*. *Food Chem.* **2005**, *90*, 151–156. [[CrossRef](#)]
24. Kimura, S. Wide distribution of the skin type I collagen alpha 3 chain in bony fish. *Comp. Biochem. Physiology. B Comp. Biochem.* **1992**, *102*, 255–260. [[CrossRef](#)]
25. Sun, L.; Hou, H.; Li, B.; Zhang, Y. Characterization of acid-and pepsin-soluble collagen extracted from the skin of Nile tilapia (*Oreochromis niloticus*). *Int. J. Biol. Macromol.* **2017**, *99*, 8–14. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, X.; Zhang, H.; Toriumi, S.; Ura, K.; Takagi, Y. Feasibility of collagens obtained from bester sturgeon *Huso huso* × *Acipenser ruthenus* for industrial use. *Aquaculture* **2020**, *529*, 735641. [[CrossRef](#)]
27. Duan, R.; Zhang, J.; Du, X.; Yao, X.; Konno, K. Properties of collagen from skin, scale and bone of carp (*Cyprinus carpio*). *Food Chem.* **2009**, *112*, 702–706. [[CrossRef](#)]
28. Han, Q.-Y.; Koyama, T.; Watabe, S.; Nagashima, Y.; Ishizaki, S. Isolation and Characterization of Collagen and Collagen Peptides with Hyaluronidase Inhibition Activity Derived from the Skin of Marlin (*Istiophoridae*). *Molecules* **2023**, *28*, 889. [[CrossRef](#)] [[PubMed](#)]
29. Wang, L.; Liang, Q.; Wang, Z.; Xu, J.; Liu, Y.; Ma, H. Preparation and characterisation of type I and V collagens from the skin of Amur sturgeon (*Acipenser schrenckii*). *Food Chem.* **2014**, *148*, 410–414. [[CrossRef](#)] [[PubMed](#)]
30. Vuorio, E.; De Crombrughe, B. The family of collagen genes. *Annu. Rev. Biochem.* **1990**, *59*, 837–872. [[CrossRef](#)] [[PubMed](#)]
31. Gauza-Włodarczyk, M.; Kubisz, L.; Włodarczyk, D. Amino acid composition in determination of collagen origin and assessment of physical factors effects. *Int. J. Biol. Macromol.* **2017**, *104*, 987–991. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, X.; Azuma, N.; Hagihara, S.; Adachi, S.; Ura, K.; Takagi, Y. Characterization of type I and II procollagen $\alpha 1$ chain in Amur sturgeon (*Acipenser schrenckii*) and comparison of their gene expression. *Gene* **2016**, *579*, 8–16. [[CrossRef](#)] [[PubMed](#)]
33. Hwang, J.-H.; Mizuta, S.; Yokoyama, Y.; Yoshinaka, R. Purification and characterization of molecular species of collagen in the skin of skate (*Raja kenoi*). *Food Chem.* **2007**, *100*, 921–925. [[CrossRef](#)]
34. Jollès, P.; Caen, J.P. Technology, Parallels between milk clotting and blood clotting: Opportunities for milk-derived products. *Trends Food Sci.* **1991**, *2*, 42–43. [[CrossRef](#)]
35. Engel, J.; Bächinger, H.P. Structure, stability and folding of the collagen triple helix. In *Collagen: Primer in Structure, Processing Assembly*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 7–33.
36. Gaill, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R. Structural comparison of cuticle and interstitial collagens from annelids living in shallow sea-water and at deep-sea hydrothermal vents. *J. Mol. Biol.* **1995**, *246*, 284–294. [[CrossRef](#)]
37. Tanaka, T.; Takahashi, K.; Adachi, K.; Ohta, H.; Yoshimura, Y.; Agawa, Y.; Sawada, Y.; Takaoka, O.; Biswas, A.K.; Takii, K. Molecular cloning and expression profiling of procollagen $\alpha 1$ (I) of cultured Pacific bluefin tuna. *Fish. Sci.* **2014**, *80*, 603–612. [[CrossRef](#)]
38. Chu, M.-L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R. Amino acid sequence of the triple-helical domain of human collagen type VI. *J. Biol. Chem.* **1988**, *263*, 18601–18606. [[CrossRef](#)] [[PubMed](#)]
39. Saito, M.; Takenouchi, Y.; Kunisaki, N.; Kimura, S. Complete primary structure of rainbow trout type I collagen consisting of $\alpha 1$ (I) $\alpha 2$ (I) $\alpha 3$ (I) heterotrimers. *Eur. J. Biochem.* **2001**, *268*, 2817–2827. [[CrossRef](#)]
40. Von der Mark, K. Structure, biosynthesis and gene regulation of collagens in cartilage and bone. In *Dynamics of Bone and Cartilage Metabolism: Principles and Clinical Applications*; Elsevier: Amsterdam, The Netherlands, 2006; pp. 3–40.
41. Bernard, M.; Yoshioka, H.; Rodriguez, E.; Van der Rest, M.; Kimura, T.; Ninomiya, Y.; Olsen, B.; Ramirez, F. Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *J. Biol. Chem.* **1988**, *263*, 17159–17166. [[CrossRef](#)]
42. Isenberg, J.S.; Frazier, W.A.; Krishna, M.C.; Wink, D.A.; Roberts, D.D. Enhancing cardiovascular dynamics by inhibition of thrombospondin-1/CD47 signaling. *Curr. Drug Targets* **2008**, *9*, 833–841. [[CrossRef](#)]
43. Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R.-N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671. [[CrossRef](#)]
44. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
45. Lin, K.; Zhang, L.; Han, X.; Xin, L.; Meng, Z.; Gong, P.; Cheng, D. Yak milk casein as potential precursor of angiotensin I-converting enzyme inhibitory peptides based on in silico proteolysis. *Food Chem.* **2018**, *254*, 340–347. [[CrossRef](#)] [[PubMed](#)]
46. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685. [[CrossRef](#)]
47. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. [[CrossRef](#)] [[PubMed](#)]

48. DeLano, W.L. The PyMOL Molecular Graphics System. 2002. Available online: <http://www.pymol.org/> (accessed on 2 January 2024).
49. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [[CrossRef](#)] [[PubMed](#)]
50. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. BIOPEP-UWM database of bioactive peptides: Current opportunities. *Int. J. Mol. Sci.* **2019**, *20*, 5978. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.