Research Article

In-depth correlation analysis demonstrates that 4-hydroxyproline at the Yaa position of Gly-Xaa-Yaa repeats dominantly stabilizes collagen triple helix

Yuki Taga * **Keisuke Tanaka, Shunji Hattori and Kazunori Mizuno**

Nippi Research Institute of Biomatrix, *520-11 Kuwabara, Toride, Ibaraki 302-0017, Japan*

Correspondence to Yuki Taga: *y-taga@nippi-inc.co.jp (Y. Taga), ktanaka@nippi-inc.co.jp (K. Tanaka), shunhatt@nippi-inc.co.jp (S. Hattori), k-mizuno@nippi-inc.co.jp (K. Mizuno)*
https://doi.org/10.1016/j.mbplus.2021.100067

Abstract

There is a general consensus that collagen stability is largely maintained by Pro and its major hydroxylated form, 4-hydroxyproline (4Hyp). However, positional difference in their stabilizing effect at the Xaa or Yaa position of collagenous Gly-Xaa-Yaa sequences has remained inconclusive. Here, we position-specifically evaluated the correlation of imino acid contents to denaturation temperature (T_d) of collagen among various vertebrate and invertebrate species, using a recently developed LC–MS methodology. 4Hyp at the Yaa position showed the highest positive correlation with T_d, followed by Pro at the Xaa position, which was even further increased by excluding invertebrates. We confirmed that Gly-Pro-4Hyp liberated after bacterial collagenase digestion was highly positively correlated with T_d. Furthermore, other tripeptides with Yaa position 4Hyp also had comparable positive correlation, excepting negative correlation of Gly-Gly-4Hyp, while tripeptides with Xaa position Pro did not. These data provide evidence that 4Hyp dominantly contributes to thermal stability of collagen depending on its sequence position, especially in vertebrates.

© **2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).**

Introduction

Collagen is the most abundant structural protein in animals. The primary structural feature of collagen is repeating Gly-Xaa-Yaa amino acid sequences where Pro and hydroxyproline (Hyp) are frequently found at the Xaa and Yaa positions, respectively. In general, most post-translational hydroxylation reaction of Pro residues occurs at the Yaa position generating 4Hyp, while 4Hyp and 3Hyp are rarely generated at the Xaa position in Gly-Xaa-Yaa and Gly-Xaa-4Hyp sequences, respectively [1–3]. Three polypeptide chains of collagen form a triple helical structure, which is important not only for the structural integrity but also for the biological functions [4].

The thermal stability of collagen arises from many different factors, such as interstrand hydrogen bonding between Gly and amino acids at the Xaa position and stereochemical restrictions by the imino acid rings of Pro and Hyp [5]. Since substitution of Pro with 4Hyp at the Yaa position increased the stability [6] and provided additional water bridges in model peptides [7], it had been thought that collagen is stabilized mainly by hydrogen bonding between the 4-hydroxyl groups and main-chain oxygens. However, Holmgren et al. [8] reported that the stability of collagen was also increased by (2*S*,4*R*)-4-fluoroproline whose side chain does not form hydrogen bonds, which indicates that water bridges do not contribute significantly to the stabilization in solution [9,10]. Vitagliano et al. [11,12] proposed "propensity-based" stabilization of Pro at the Xaa position [Cγ-endo (down-puckering) conformation of the imino acid ring] and 4Hyp at the Yaa position [Cγ-exo (up-puckering)]. This propensity-based hypothesis well explains the positional difference in stabilizing effects of Hyp and Pro derivatives

0022-2836/© 2021 The Authors. Published by Elsevier B.V.This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

observed for model peptides [5]. However, there is little data to support this hypothesis in real collagen molecules.

Previous studies have shown the particular importance of imino acids for collagen stability by comparative analysis of collagens from various species. Piez & Gross [13] reported that the collagen thermostability estimated by shrinkage temperature well correlated with the total content of imino acids rather than Pro and Hyp alone. In contrast, Burjanadze [14,15] reported that Hyp showed the best correlation with denaturation temperature (T_d) rather than the sum of imino acids, and suggested that the stabilizing effect is mostly exerted by 4Hyp located at the Yaa position, not by 4Hyp and 3Hyp at the Xaa position. However, these studies were based on comparison of data collected from various literature sources in which Pro and Hyp were analyzed without information of the Xaa or Yaa position in the repeat sequences.

Mass spectrometric approaches with protease digestion have been increasingly used for comprehensive identification of prolyl hydroxylation sites in collagen [16–21], but it is difficult to totally evaluate the degree of the modification with discrimination of the sequence position. Recently, we developed a novel method to quantitatively analyze positional distribution of Pro and Hyp by LC–MS with partial acid hydrolysis [3]. The method enables determination of the imino acid contents as residues/1000 amino acid residues with discrimination of the position (Xaa or Yaa) and hydroxylation type (Pro, 4Hyp, or 3Hyp). In the present study, we aimed to clarify the real contribution of imino acids to the stability of collagen. Collagens from 24 animal species, including 18 vertebrates and 6 invertebrates, were simultaneously analyzed, and the correlation of the contents of Pro and Hyp with T_d was position-specifically evaluated. Furthermore, LC–MS analysis of bacterial collagenase-liberated tripeptides was performed to verify obtained results.

Results

Correlation of total imino acids with collagen thermal stability

Circular dichroism analysis was used to determine T_d of the collagen samples, which ranged from 20.7 °C to 43.1 °C (Supplementary Fig. 1 and Supplementary Table 1). T_d was relatively low for fishes and other aquatic animals, which was consistent with the established understanding that T_d of collagen depends on environmental temperature [22]. The purity of collagen samples was validated by their amino acid composition determined after acid hydrolysis, especially by the Gly content (approximately more than 330 residues/1000 residues; Supplementary Table 2), which agreed with the fact that Gly generally accounts for approximately 33% of total amino acids in collagen that consists of Gly-Xaa-Yaa repeats.

We first evaluated the correlation of total Pro, total Hyp, and their sum to T_d using the above amino acid analysis. Positive correlation with T_d was observed for Pro ($r = 0.482$, $p = 0.017$; Fig. 1A) and the sum of Pro and Hyp ($r = 0.823$, $p = 7.75E{-}07$; Fig. 1C). Although total Hyp did not show significant correlation among all species due to a large deviation of earthworm, it became significant by limiting the data to vertebrates ($r = 0.934$, $p = 1.46E{-}08$; Fig. 1B). Under this traditional approach that has been adopted in previous studies [13–15], the sum of Pro and Hyp appeared to be most correlated with the thermal stability of collagen.

Correlation of position-specific imino acids with collagen thermal stability

We further investigated these correlations by analyzing Pro and Hyp via discrimination of their position in the Gly-Xaa-Yaa repeat sequences using our LC–MS methodology [3]. While large proportions of Pro and 4Hyp were present at the Xaa and Yaa positions, respectively, small amounts of 4Hyp/3Hyp at the Xaa position and Pro at the Yaa position were also detected in all animal species (Supplementary Table 3), consistent with our understanding of collagen prolyl hydroxylation [1–3]. Total content of imino acids at respective positions largely varied among animal species (70.2–112.8 residues/1000 residues at the Xaa position and 67.8–110.5 residues/1000 residues at the Yaa position). Accurate quantitation of 3Hyp was difficult for some species, especially fishes, due to its inherently low levels, whereas the 3Hyp content was relatively high in invertebrates.

Among the analyzed imino acids, 4Hyp at the Yaa position showed strong positive correlation with T_d ($r = 0.887$, $p = 7.83E{-}09$; Fig. 2C), and Pro at the Xaa position also correlated with T_d ($r = 0.516$, $p = 0.010$; Fig. 2D). Multiple regression analysis indicated that Yaa position 4Hyp has higher contribution to T_d than Xaa position Pro (Supplementary Table 4). These correlations were stronger than those of Pro and Hyp estimated as total contents (Fig. 1), respectively, although levels in several invertebrate species still deviated from the correlation lines. In contrast, 4Hyp at the Xaa position did not show such positive correlation (Fig. 2A). The amount of the unusual 4Hyp was especially high in earthworm collagen, as reported previously [3]. When invertebrate collagens were removed from this dataset, Xaa position 4Hyp actually negatively correlated with T_d ($r = -0.666$, $p = 2.56E{-}03$). Xaa position 3Hyp and Yaa position Pro did not exhibit any correlation (Fig. 2B and E).

The range of variation in prolyl 4-hydroxylation at the Yaa position among analyzed collagens was small (77.8%–95.2%; Supplementary Table 5). A

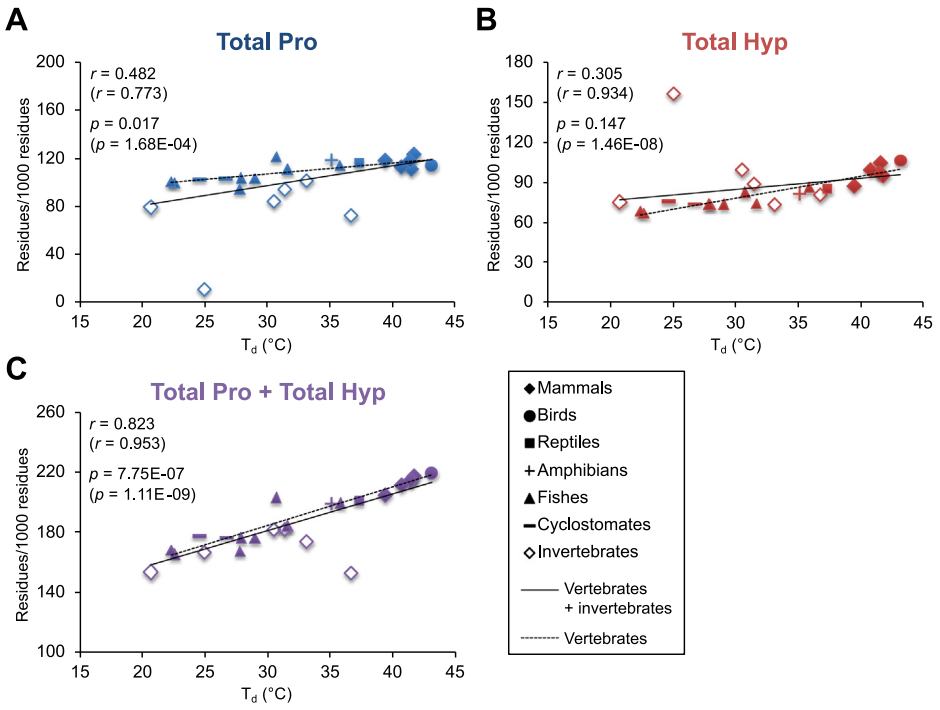

Fig. 1. Correlation analysis between total imino acid contents and T_d in collagens from various species. Correlation with T_d was analyzed for total contents of (A) Pro, (B) Hyp, and (C) Pro + Hyp. Pearson's correlation coefficient (*r*) and statistical significance (*p*) are shown. Values in parentheses are calculated by excluding invertebrates.

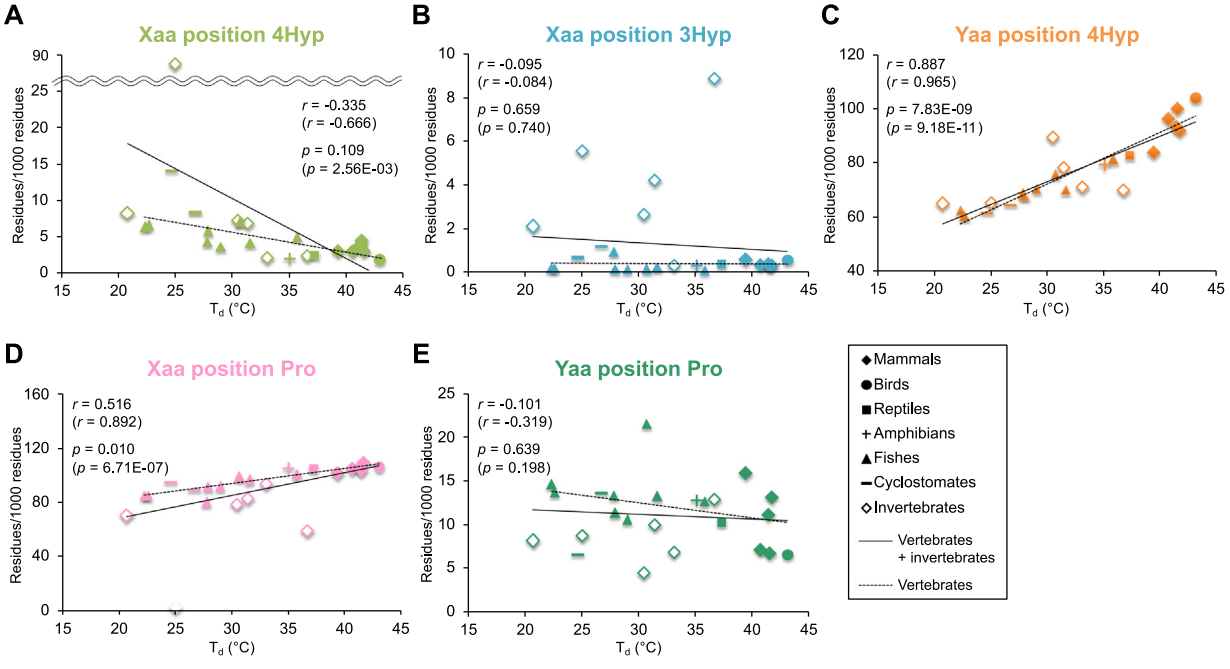

Fig. 2. Correlation analysis between position-specific imino acid contents and T_d in collagens from various species. Correlation with T_d was analyzed for (A and C) 4Hyp, (B) 3Hyp, and (D and E) Pro at the (A, B, and D) Xaa or (C and E) Yaa positions. Pearson's correlation coefficient (*r*) and statistical significance (*p*) are shown. Values in parentheses are calculated by excluding invertebrates.

significant positive correlation with T_d was detected for the prolyl 4-hydroxylation rate at the Yaa position (Supplementary Fig. 2C), but it was weak relative to that for the content of Yaa position 4Hyp. The rate of prolyl 4- and 3-hydroxylation at the Xaa position showed similar tendency with the content of 4Hyp and 3Hyp at the position, respectively, in the correlation relative to T_d (Supplementary Fig. 2A and B).

Correlation of collagenase-liberated tripeptides with collagen thermal stability

To verify these findings, correlation of imino acids with T_d was evaluated by another approach employing LC–MS detection of tripeptides liberated by bacterial collagenase digestion. Gly-Pro-4Hyp and other major tripeptide sequences were selected for Gly-Xaa-4Hyp tripeptides (Xaa = Ala, Glu, and Leu) and Gly-Pro-Yaa tripeptides (Yaa = Ala, Arg, and Gln) [23]. In addition, we analyzed tripeptide sequences with Xaa position 4Hyp, including Gly-4Hyp-Ala and Gly-4Hyp-Gly, which were previously identified in human collagens [18,24]. To discriminate identical molecular weight tripeptides, such as Gly-Ala-4Hyp and Gly-4Hyp-Ala, we set multiple reaction monitoring (MRM) channels with specific fragment ions at *m/z* 132 and 143 for Gly-Xaa-4Hyp and Gly-4Hyp-Yaa, respectively (Supplementary Fig. 3). Similarly, a fragment ion at *m/z* 127 was used for specific detection of Gly-Pro-Yaa.

Gly-Xaa-4Hyp and Gly-Pro-Yaa tripeptides were abundantly detected after collagenase digestion, consistent with their occurrence in fibrillar collagens [23], while the amount of Gly-4Hyp-Yaa tripeptides was relatively low compared to those major tripeptides, except in earthworm (Supplementary Table 6). It was confirmed that the amount of collagenase-liberated tripeptides matched with the theoretical number of tripeptide units for bovine type I collagen (Supplementary Fig. 4). There were interesting differences in the relationships of the tripeptide sequences to T_d depending on the position of Pro/4Hyp and adjacent amino acids (Fig. 3 and Supplementary Fig. 5). Clear positive correlation was observed for Gly-Pro-4Hyp (r = 0.698, p = 1.49E−04; Fig. 3A), although some invertebrates showed low contents of this tripeptide. The sum of Gly-Xaa-4Hyp tripeptides was also positively correlated with T_d (r = 0.763, p = 1.44E−05; Fig. 3B), which was comparable to that of Gly-Pro-4Hyp. In particular, Gly-Glu-4Hyp showed the highest positive correlation among analyzed tripeptides (r = 0.844, p = 2.23E−07; Supplementary Fig. 5B). In contrast, no significant correlation was observed for the sum of Gly-Pro-Yaa tripeptides (Fig. 3C). Only Gly-Pro-Arg exhibited positive correlation with a slight statistical significance among vertebrates (r = 0.475, p = 0.046; Supplementary Fig. 5E). Although a tendency toward negative correlation was observed for Xaa position 4Hyp (Fig. 2A), Gly-4Hyp-Yaa tripeptides were not correlated with T_d even among vertebrates (Fig. 3D). On the other hand, an irregular tripeptide with Yaa position 4Hyp, Gly-Gly-4Hyp, showed negative correlation with T_d (r = −0.721, p = 7.06E−05; Fig. 3E).

Discussion

In the present study, we simultaneously analyzed collagens from 24 species, encompassing both vertebrates and invertebrates, and enabled position-specific evaluation of the relationship of imino acid contents to T_d by two approaches employing partial acid hydrolysis or bacterial collagenase digestion. Both experiments indicated that 4Hyp had the highest correlation with T_d only when at the Yaa position, demonstrating the dominating role of Yaa position 4Hyp for stabilization of the collagen triple helix. Its particularly significant contribution is confirmed by the observation that 4Hyp at the Yaa position correlated better with T_d than the total amount of Hyp that additionally includes 4Hyp/3Hyp at the Xaa position. In addition, the stabilizing effect of Yaa position 4Hyp was exemplified in the collagens from two octopus species (*Octopus sinensis* and *Enteroctopus dofleini*), which showed a 10 °C difference in T_d (Supplementary Table 1). The amino acid composition was almost identical between the octopus collagens (Supplementary Table 2), but 4Hyp assigned to the Yaa position was shown to be markedly high in *O. sinensis* by both analytical methods (Supplementary Tables 3 and 6). This data supports the notion that T_d primarily depends on the content of 4Hyp at the Yaa position, although contribution of other imino acids showing increases in *O. sinensis*, such as Xaa position Pro, cannot be ruled out. Based on the small variation in prolyl 4-hydroxylation at the Yaa position among animals, we can consider that the large variation in the abundance of Yaa position 4Hyp is mainly programmed by the frequency of Pro at this position in the primary amino acid sequence, not only by the degree of prolyl 4-hydroxylation. Deficiency in prolyl 4-hydroxylase caused significant reduction in T_d of collagen in tissue [25], and inhibition of the enzyme impaired proper procollagen production from cells [26]. On the other hand, in our previous study, the degree of prolyl 4-hydroxylation remained definitely constant with aging [27]. These observations indicate the essential role of 4Hyp in the body.

Host-guest peptides $(Gly\text{-}Pro\text{-}4Hyp)_3\text{-}Gly\text{-}Xaa\text{-}Yaa\text{-}(Gly\text{-}Pro\text{-}4Hyp)_4$ that form a triple helical conformation have been used to investigate the effect of the guest peptide sequence, Gly-Xaa-Yaa, on the stability of the triple helix [23,28,29]. In the modeling experiments, Pro and 4Hyp are predicted to be the most stabilizing residues at the Xaa and Yaa positions, respectively, among 20 amino acids representing a wide range of stability; there-

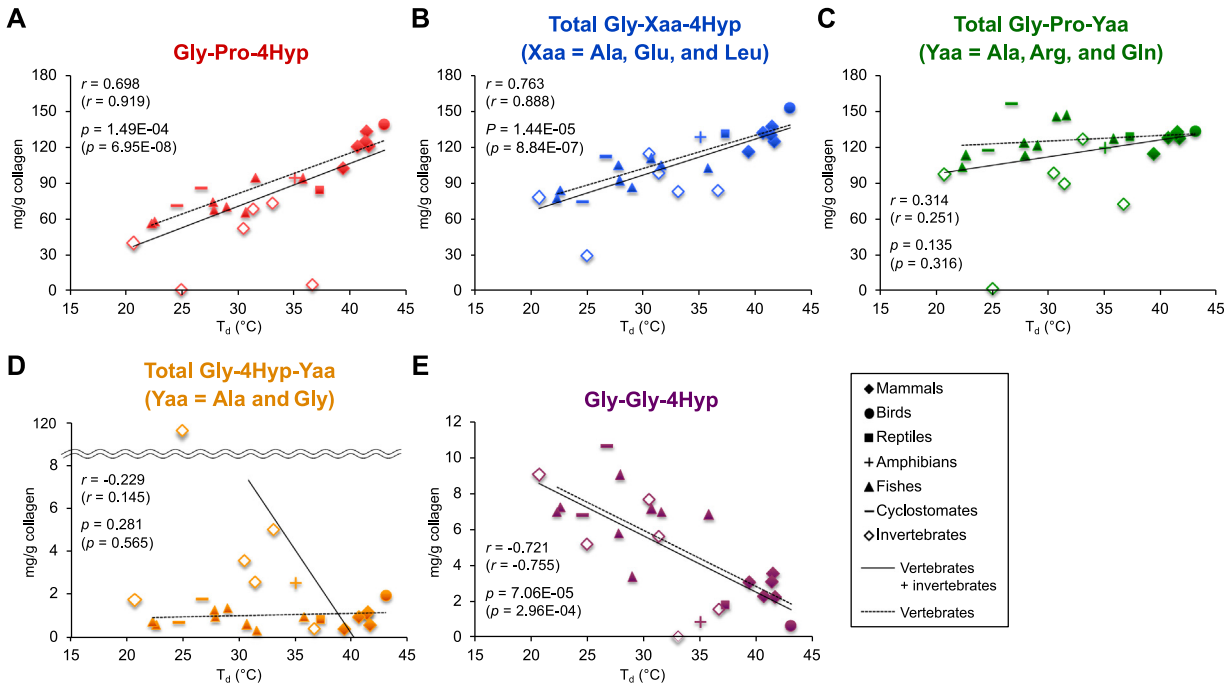

Fig. 3. Correlation analysis between amounts of collagenase-liberated tripeptides and T_d in collagens from various species. Correlation with T_d was analyzed for (A) Gly-Pro-4Hyp, (B) all Gly-Xaa-4Hyp tripeptides (Gly-Ala-4Hyp + Gly-Glu-4Hyp + Gly-Leu-4Hyp), (C) all Gly-Pro-Yaa tripeptides (Gly-Pro-Ala + Gly-Pro-Arg + Gly-Pro-Gln), (D) all Gly-4Hyp-Yaa tripeptides (Gly-4Hyp-Ala + Gly-4Hyp-Gly), and (E) Gly-Gly-4Hyp. Pearson's correlation coefficient (*r*) and statistical significance (*p*) are shown. Values in parentheses are calculated by excluding invertebrates.

fore Gly-Pro-4Hyp is predicted to be the most stabilizing triplet sequence [23]. The frequency of occurrence of amino acids at the respective positions in collagen sequences is correlated with thermal stability of the host-guest peptides with guest triplets Gly-Xaa-4Hyp or Gly-Pro-Yaa [23]. Consistent with this, we confirmed that frequently occurring Gly-X-4Hyp sequences, including Gly-Pro-4Hyp (32.9%), Gly-Glu-4Hyp (13.0%), and Gly-Leu-4Hyp (7.8%), except for Gly-Ala-4Hyp (11.1%) [23], were positively correlated with T_d among collagens from various species. Gly-Glu-4Hyp showed particularly high correlation, surpassing that of Gly-Pro-4Hyp, despite its relatively low abundance among the analyzed Gly-Xaa-4Hyp tripeptides. Glu with a negative charge is predominantly located at the Xaa position, while Arg and Lys with a positive charge are predominantly located at the Yaa position [23]. Glu may critically participate in the stabilization of collagen by the formation of interstrand electrostatic interactions via the charged side chain [30]. A previous study by Burjanadze [31] suggested that the Gly-Pro-4Hyp sequence is the dominating factor influencing collagen thermostability based on correlation analysis using probable numbers of Gly-Pro-4Hyp and Gly-Xaa-4Hyp sequences in animal collagens calculated from amino acid composition. However, our data provide substantial evidence that Gly-Xaa-4Hyp sequences also play a role just as important as Gly-Pro-4Hyp in collagen stability.

In contrast to 4Hyp at the Yaa position, Pro at the Xaa position showed somewhat inconsistent results between the two analytical methods used. While the second highest positive correlation to T_d was observed for Xaa position Pro, there was no correlation for major Gly-Pro-Yaa tripeptides, excepting only slight correlation in vertebrates for Gly-Pro-Arg reported to confer high stability comparable to Gly-Pro-4Hyp in host-guest experiments [29]. As we did not analyze all Gly-Pro-Yaa tripeptides, we cannot conclude actual contribution of Pro at the Xaa position. However, our data suggest that the effect of Pro on the stability of collagen molecules is limited, and that the significant positive correlation observed for Xaa position Pro largely reflected the stabilizing ability of 4Hyp in the Gly-Pro-4Hyp sequence.

Remarkably, all the positive correlations to T_d were increased by excluding invertebrate collagens from the analysis. The very high correlation observed for Yaa position 4Hyp (r = 0.965, p = 9.18E−11; Fig. 2C), Gly-Pro-4Hyp (r = 0.919, p = 6.95E−08; Fig. 3A), and total Gly-Xaa-4Hyp (r = 0.888, p = 8.84E−07; Fig. 3B) among vertebrates suggests that contributions of other factors to T_d are little or essentially common

in vertebrate collagens. Invertebrates tended to be below those correlation lines. Other factors, such as glycosylation [32], charged amino acids [33], and disulfide bonding [34], reported to be involved in the stabilization of the triple helical structure of invertebrate collagens, may compensate for the insufficiency of 4Hyp at the Yaa position.

Although it was expected that Gly-4Hyp-Yaa tripeptides would negatively correlate with T_d based on the results of 4Hyp assigned to the Xaa position, no correlation was detected for Gly-4Hyp-Ala or Gly-4Hyp-Gly. Since the method used to quantitate the position-specific contents of Pro/Hyp judges the location of the imino acids from the positional relationship to Gly, 4Hyp at the Yaa position of Gly-Gly-Yaa-Gly sequences was counted as occupying both the Xaa and Yaa positions [3]. The amount of Gly-Gly-4Hyp was relatively high compared to the two Gly-4Hyp-Yaa tripeptides analyzed here, suggesting that the irregular sequence with Yaa position 4Hyp significantly affected the estimated values of Xaa position 4Hyp. Indeed, analysis of collagenase-liberated tripeptides clarified negative correlation of Gly-Gly-4Hyp similar to Xaa position 4Hyp. Gly-Gly-Yaa sequences are reported to destabilize the triple helical conformation [28]. The negative correlation detected for Gly-Gly-4Hyp would be due to reflecting the total abundance of Gly-Gly-Yaa sequences, not Gly-Gly-4Hyp alone. Considering the higher content of Gly residues (>340 residues/1000 residues), fishes, cyclostomates, and invertebrates likely have non-negligible amounts of Gly-Gly-Yaa sequences in addition to Gly-Gly-4Hyp. These irregular sequences may substantially contribute to the relatively low T_d of these species.

There was no meaningful significant correlation with T_d for imino acids other than Yaa position 4Hyp and Xaa position Pro. However, the remaining unhydroxylated Pro at the Yaa position (~5%–20%; Supplementary Table 5) may enable flexible regulation of the thermal stability of collagen. In addition, although the role of 4Hyp at the Xaa position is unclear, the diverse range of its abundance among animals (Supplementary Table 6) suggests that this unusual modification contributes to the adaptation to subtle changes in environmental temperature, such as seasonal ones. According to the propensity-based hypothesis, 4Hyp preferring up-puckering potentially destabilizes the triple helix when it locates at the Xaa position, while 3Hyp preferring down-puckering potentially stabilizes it [5]. We found no correlation between Xaa position 4Hyp/3Hyp and T_d; however, since the abundance of these minor Hyp residues appears to be too low to perform the correlation analysis, we cannot rule out the possibility that prolyl hydroxylation at the Xaa position has some regulatory role in the thermostability of collagen.

In conclusion, we provide evidence that 4Hyp at the Yaa position of Gly-Xaa-Yaa repeat sequences dominantly contributes to the thermal stability of collagen, especially in vertebrates. To the best of our knowledge, this is the first study that comprehensively evaluates the relationship of Pro/Hyp contents and T_d of real collagen molecules with discrimination of their sequence positions. We basically used type I (-like) collagen extracted from skin with salting-out purification to remove several other collagen types, such as types IV and V (type III collagen would be contained to some extent). However, the details about invertebrate collagens are often unknown, and the relationship between imino acid contents and T_d may be largely different depending on the type of collagen. Further experiments using various invertebrate collagens and other collagen types may provide further understanding of the importance of prolyl hydroxylation in the regulation of thermal stability of the collagen triple helix.

Experimental procedures

Materials and reagents

Pepsin and 3-aminopyridyl-*N*-hydroxysuccinimidyl carbamate (APDS) were purchased from Wako Chemicals (Osaka, Japan), and recombinant collagenase from *Grimontia hollisae* was procured from Nippi (Tokyo, Japan). Gly-Pro, Pro-Gly, 4Hyp-Gly, Gly-Ala-4Hyp, Gly-Pro-4Hyp, Gly-Pro-Ala, and Gly-Pro-Arg were purchased from Bachem (Bubendorf, Switzerland), Gly-4Hyp was purchased from Santa Cruz Biotechnology (Santa Cruz, CA), and other peptides were custom synthesized by AnyGen (Gwangju, Korea).

We used collagens previously extracted/purified from animal tissues and stored in our laboratory (Supplementary Table 1; N = 1 for respective species). Almost all of the collagen samples were from skin, except for those from earthworm (cuticle), sea cucumber (body wall), and sea anemone (whole body). Since the extraction and purification methods were sometimes different among the collagen samples, we further performed digestion with pepsin (0.1 mg/mL in 0.5 M acetic acid at 4 °C for 16 h), salt precipitation (0.7 M NaCl in 0.5 M acetic acid on ice for 3 h), and isoelectric precipitation (0.1 M Tris-HCl, pH 8.0, 4 °C for 16 h; except for earthworm collagen, which was precipitated with 33% ethanol on ice for 1 h).

Circular dichroism determination of T_d

Collagen samples were diluted to a concentration of 0.1 mg/mL in 5 mM acetic acid and analyzed using a J-805 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan). T_d was defined as

the temperature of the midpoint of circular dichroism ellipticity at 221 nm between 10 °C and 40 °C or 20 °C and 50 °C with a constant increase in temperature (0.25 °C/min).

Determination of amino acid composition

Gas-phase acid hydrolysis of collagen samples was performed with 6 N HCl/1% phenol at 110 °C for 20 h under N_2, as described previously [3]. The acid hydrolysate was analyzed using an L-8900 amino acid analyzer (Hitachi, Tokyo, Japan) with citrate buffer and a sodium chloride gradient. Amino acids in the eluate were monitored by postcolumn reaction with ninhydrin, and the amino acid composition was determined as the number of residues/1000 amino acid residues.

Position-specific quantitation of Pro and Hyp

Liquid-phase partial acid hydrolysis [HCl/TFA/water (2:1:1, v/v) at 55 °C for 16 h] of collagen samples and subsequent derivatization with APDS were performed as described previously [3]. The reactant was subjected to LC–MS analysis on a 3200 QTRAP mass spectrometer (AB Sciex, Foster City, CA) coupled to an Agilent 1200 Series HPLC system (Agilent Technologies, Palo Alto, CA). APDS-derivatized dipeptides, including Gly-Pro, Gly-4Hyp, Gly-3Hyp, Pro-Gly, and 4Hyp-Gly, were separated on a Hypercarb column (3 μm particle size, L × I.D. 100 mm × 2.1 mm; Thermo Fisher Scientific, Waltham, MA) and detected in MRM mode [3]. The position-specific contents of Pro and Hyp were calculated from the ratio between Gly-Pro/Pro-Gly and Gly-4Hyp/Gly-3Hyp/4Hyp-Gly, respectively, in combination with total Pro and total Hyp determined by amino acid analysis.

Quantitation of tripeptides after bacterial collagenase digestion

Collagen samples were digested with *Grimontia* collagenase in 100 mM Tris-HCl/5 mM $CaCl_2$ (pH 7.5) at 37 °C for 16 h after heat denaturation at 60 °C for 30 min. Liberated tripeptides were quantitated by LC–MS analysis in MRM mode following chromatographic separation using an Ascentis Express F5 HPLC column (5 μm particle size, L × I.D. 250 mm × 4.6 mm; Supelco, Bellefonte, PA), as described previously [35]. The following MRM transitions were monitored: *m/z* 260.1 → 132.1 (Gly-Ala-4Hyp), *m/z* 318.2 → 132.0 (Gly-Glu-4Hyp), *m/z* 302.2 → 132.1 (Gly-Leu-4Hyp), *m/z* 286.0 → 127.2 (Gly-Pro-4Hyp), *m/z* 244.1 → 127.2 (Gly-Pro-Ala), *m/z* 329.2 → 127.1 (Gly-Pro-Arg), *m/z* 301.0 → 127.1 (Gly-Pro-Gln), *m/z* 260.1 → 143.1 (Gly-4Hyp-Ala), *m/z* 246.2 → 143.1 (Gly-4Hyp-Gly), and *m/z* 246.2 → 132.1 (Gly-Gly-4Hyp). The amounts of liberated tripeptides relative to original collagen samples were calculated based on external calibration curves prepared using synthetic peptide standards.

Statistical analysis

The correlation coefficient (*r*) of imino acid contents or amounts of collagenase-liberated tripeptides to T_d were analyzed by Pearson's correlation test with Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA). Multiple regression analysis was performed to evaluate the contribution of Xaa position Pro and Yaa position 4Hyp (independent variable) to T_d (dependent variable) using InStat version 3.0b for Macintosh (GraphPad Software, San Diego, CA). There was no multicollinearity between the variables. *p* values less than 0.05 were considered statistically significant.

CRediT authorship contribution statement

Yuki Taga: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing - original draft. **Keisuke Tanaka**: Resources, Writing - review & editing. **Shunji Hattori**: Resources, Writing - review & editing. **Kazunori Mizuno**: Supervision, 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mbplus.2021.100067.

Received 1 February 2021;
Accepted 19 April 2021;
Available online 7 May 2021

Keywords:
Collagen;
Hydroxyproline;
Denaturation temperature;
Thermal stability

† Present address: Helmholtz Institute for pharmaceutical sciences Campus E8 1, 66123 Saarbrücken, Germany.

Abbreviations:
Hyp, hydroxyproline; T_d, denaturation temperature; MRM, multiple reaction monitoring; APDS, 3-aminopyridyl-*N*-hydroxysuccinimidyl carbamate

References

1. Gorres, K.L., Raines, R.T., (2010). Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.*, **45** (2), 106–124.
2. Hudson, D.M., Eyre, D.R., (2013). Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification?. *Connect. Tissue Res.*, **54** (4–5), 245–251.
3. Taga, Y., Kusubata, M., Mizuno, K., (2020). Quantitative analysis of the positional distribution of hydroxyproline in collagenous Gly-Xaa-Yaa sequences by LC-MS with partial acid hydrolysis and precolumn derivatization. *Anal. Chem.*, **92** (12), 8427–8434.
4. Brodsky, B., Ramshaw, J.A., (1997). The collagen triple-helix structure. *Matrix Biol.*, **15** (8–9), 545–554.
5. Bella, J., (2016). Collagen structure: new tricks from a very old dog. *Biochem. J.*, **473** (8), 1001–1025.
6. Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., Prockop, D.J., (1973). Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. *Biochim. Biophys. Acta*, **303** (1), 198–202.
7. Bella, J., Eaton, M., Brodsky, B., Berman, H.M., (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science*, **266** (5182), 75–81.
8. Holmgren, S.K., Taylor, K.M., Bretscher, L.E., Raines, R. T., (1998). Code for collagen's stability deciphered. *Nature*, **392** (6677), 666–667.
9. Engel, J., Prockop, D.J., (1998). Does bound water contribute to the stability of collagen?. *Matrix Biol.*, **17** (8–9), 679–680.
10. Shoulders, M.D., Raines, R.T., (2009). Collagen structure and stability. *Annu. Rev. Biochem.*, **78**, 929–958.
11. Vitagliano, L., Berisio, R., Mazzarella, L., Zagari, A., (2001). Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers*, **58** (5), 459–464.
12. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., Zagari, A., (2001). Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability. *Protein Sci.*, **10** (12), 2627–2632.
13. Piez, K.A., Gross, J., (1960). The amino acid composition of some fish collagens: the relation between composition and structure. *J. Biol. Chem.*, **235**, 995–998.
14. Burjanadze, T.V., (1979). Hydroxyproline content and location in relation to collagen thermal stability. *Biopolymers*, **18** (4), 931–938.
15. Burjanadze, T.V., (1982). Evidence for the role of 4-hydroxyproline localized in the third position of the triplet (Gly-X-Y) in adaptational changes of thermostability of a collagen molecule and collagen fibrils. *Biopolymers*, **21** (8), 1489–1501.
16. Montgomery, H., Rustogi, N., Hadjisavvas, A., Tanaka, K., Kyriacou, K., Sutton, C.W., (2012). Proteomic profiling of breast tissue collagens and site-specific characterization of hydroxyproline residues of collagen alpha-1-(I). *J. Proteome Res.*, **11** (12), 5890–5902.
17. Ma, F., Sun, R., Tremmel, D.M., Sackett, S.D., Odorico, J., Li, L., (2018). Large-scale differentiation and site specific discrimination of hydroxyproline isomers by electron transfer/higher-energy collision dissociation (EThcD) mass spectrometry. *Anal. Chem.*, **90** (9), 5857–5864.
18. van Huizen, N.A., Burgers, P.C., Saintmont, F., Brocorens, P., Gerbaux, P., Stingl, C., Dekker, L.J.M., IJzermans, J.N. M., Luider, T.M., (2019). Identification of 4-hydroxyproline at the Xaa position in collagen by mass spectrometry. *J. Proteome Res.*, **18** (5), 2045–2051.
19. Merl-Pham, J., Basak, T., Knuppel, L., Ramanujam, D., Athanason, M., Behr, J., Engelhardt, S., Eickelberg, O., Hauck, S.M., Vanacore, R., Staab-Weijnitz, C.A., (2019). Quantitative proteomic profiling of extracellular matrix and site-specific collagen post-translational modifications in an in vitro model of lung fibrosis. *Matrix Biol. Plus*, **1**, 100005
20. Cabral, W.A., Fratzl-Zelman, N., Weis, M., Perosky, J.E., Alimasa, A., Harris, R., Kang, H., Makareeva, E., Barnes, A.M., Roschger, P., Leikin, S., Klaushofer, K., Forlino, A., Backlund, P.S., Eyre, D.R., Kozloff, K.M., Marini, J.C., (2020). Substitution of murine type I collagen A1 3-hydroxylation site alters matrix structure but does not recapitulate osteogenesis imperfecta bone dysplasia. *Matrix Biol.*, **90**, 20–39.
21. Tonelli, F., Cotti, S., Leoni, L., Besio, R., Gioia, R., Marchese, L., Giorgetti, S., Villani, S., Gistelinck, C., Wagener, R., Kobbe, B., Fiedler, I.A.K., Larionova, D., Busse, B., Eyre, D., Rossi, A., Witten, P.E., Forlino, A., (2020). Crtap and p3h1 knock out zebrafish support defective collagen chaperoning as the cause of their osteogenesis imperfecta phenotype. *Matrix Biol.*, **90**, 40–60.
22. Rigby, B.J., (1968). Amino-acid composition and thermal stability of the skin collagen of the Antarctic ice-fish. *Nature*, **219** (5150), 166–167.
23. Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A., Brodsky, B., (2000). Amino acid propensities for the collagen triple-helix. *Biochemistry*, **39** (48), 14960–14967.
24. Kimura, T., Cheah, K.S., Chan, S.D., Lui, V.C., Mattei, M. G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y., Olsen, B.R., (1989). The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and genomic DNA reveals characteristics of a fibrillar collagen with differences in genomic organization. *J. Biol. Chem.*, **264** (23), 13910–13916.
25. Aro, E., Salo, A.M., Khatri, R., Finnila, M., Miinalainen, I., Sormunen, R., Pakkanen, O., Holster, T., Soininen, R., Prein, C., Clausen-Schaumann, H., Aszodi, A., Tuukkanen, J., Kivirikko, K.I., Schipani, E., Myllyharju, J., (2015). Severe extracellular matrix abnormalities and chondrodysplasia in mice lacking collagen prolyl 4-hydroxylase isoenzyme II in combination with a reduced amount of isoenzyme I. *J. Biol. Chem.*, **290** (27), 16964–16978.
26. Sasaki, T., Majamaa, K., Uitto, J., (1987). Reduction of collagen production in keloid fibroblast cultures by ethyl-3,4-dihydroxybenzoate. Inhibition of prolyl hydroxylase activity as a mechanism of action. *J. Biol. Chem.*, **262** (19), 9397–9403.
27. Taga, Y., Kusubata, M., Ogawa-Goto, K., Hattori, S., (2016). Developmental stage-dependent regulation of prolyl 3-hydroxylation in tendon type I collagen. *J. Biol. Chem.*, **291** (2), 837–847.
28. Shah, N.K., Sharma, M., Kirkpatrick, A., Ramshaw, J.A., Brodsky, B., (1997). Gly-Gly-containing triplets of low stability adjacent to a type III collagen epitope. *Biochemistry*, **36** (19), 5878–5883.

29. Yang, W., Chan, V.C., Kirkpatrick, A., Ramshaw, J.A., Brodsky, B., (1997). Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. *J. Biol. Chem.*, **272** (46), 28837–28840.
30. Gauba, V., Hartgerink, J.D., (2007). Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.*, **129** (9), 2683–2690.
31. Burjanadze, T.V., (2000). New analysis of the phylogenetic change of collagen thermostability. *Biopolymers*, **53** (6), 523–528.
32. Bann, J.G., Peyton, D.H., Bachinger, H.P., (2000). Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett.*, **473** (2), 237–240.
33. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., Brodsky, B., (2007). Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.*, **282** (41), 29757–29765.
34. McBride, O.W., Harrington, W.F., (1967). Helix-coil transition in collagen. Evidence for a single-stranded triple helix. *Biochemistry*, **6** (5), 1499–1514.
35. Taga, Y., Iwasaki, Y., Shigemura, Y., Mizuno, K., (2019). Improved in vivo tracking of orally administered collagen hydrolysate using stable isotope labeling and LC-MS techniques. *J. Agric. Food Chem.*, **67** (16), 4671–4678.