Biochemical and Biophysical Research Communications 422 (2012) 5–7

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Mini Review

Why collagens best survived in fossils? Clues from amino acid thermal stability

Shuang-Yin Wang [a,b], Enrico Cappellini [c], Hong-Yu Zhang [a,b,*]

[a] National Key Laboratory of Crop Genetic Improvement, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China
[b] Center for Bioinformatics, Huazhong Agricultural University, Wuhan 430070, PR China
[c] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 17 April 2012
Available online 30 April 2012

Keywords:
Collagen
Amino acid
Thermal stability
Molecular paleontology

ABSTRACT

Explaining why type I collagens are preferentially preserved in the geological time scale remains a challenge. Several pieces of evidence indicate that its rich content in the bone and its unique, stable structure played key roles in its preservation. By considering the distinct thermal stability of amino acids, we reveal that the elevated abundance of thermostable amino acid residues in type I collagens also contribute to its survival.

© 2012 Elsevier Inc. All rights reserved.

Molecular paleontology is ranked by *Science* Magazine as one of the top 10 scientific advances since the dawn of the new millennium [1]. The information contained in ancient molecules not only make it possible to revisit the prehistoric world and address questions of historic, evolutionary and taxonomic significance, such information also provides a new perspective for materials science [2] and clinical medicine [3].

An unexpected finding from the study of dinosaur fossil protein residues was that collagens had the potential for preservation for millions of years. In recent years, Schweitzer and his colleagues extracted and sequenced some type I collagen peptides (except for one peptide belonging to type II collagen) from 68-million-years-old fossils of *Tyrannosaurus rex* singly [4] and 80-million-years-old fossils of *Brachylophosaurus canadensis* with independent replication in multiple labs [5]. In addition, they provided several lines of evidence to support the reliability of the results. Even though osteocalcin, elastin, laminin, and hemoglobin were also detected using in situ immunohistochemical analyses, none of these proteins could be determined by mass spectrometry in any of the experiments [5,6]. In addition, collagens were also recovered with overwhelming abundance and coverage in a recent study about surviving proteins in Pleistocene mammoth femur fossils [7]. Therefore, the intriguing question to ask is why type I collagens are preferentially preserved in the geological time scale.

A straightforward explanation to this issue is that rich content [8] and limited solubility [9,10] both provide excellent prerequisites for the survival of the type I collagen. The bone is a composite material comprising an intimate association of protein fibers and mineral crystals. The most abundant proteins within the bone consist of type I collagens (approximately 90% by weight of the protein) [8]. In addition, collagens can only be brought into the solution by acid hydrolysis or partial enzymatic digestion [9,10]. Moreover, when a collagen fibril is surrounded by or absorbed in a mineral matrix, the collagen molecule is resistant to extraction by neutral salts or mild acid solutions and exempted from degradation by collagenases [9,10].

Further explanation comes from the fact that the triple-helical arrangement and intramolecular and intermolecular cross-links confer stability upon the ubiquitous structural molecule [11,12]. Collagen fibers are tough bundles of type I collagens. Monomeric type I collagen consists of three intertwined chains that form a triple helix by numerous hydrogen bonds. Every third amino acid residue in the sequence of a chain is Gly, leading to a repeating triplet. This repeating triplet is generally Gly-X_{aa}-Hyp (Hyp, hydroxyproline) or Gly-Pro-X_{aa}, where X_{aa} may be any other residues. The high abundance of Gly and Hyp in type I collagen is important in terms of ensuring the stabilization of the ubiquitous structure; it allows a very close association between both monomers and chains and facilitates hydrogen bonding and the formation of intermolecular cross-links [11,12].

San Antonio et al. provided a deeper insight into the issue by mapping the dinosaur fossil-derived peptides to the high-resolution model of type I collagen fibrils of extant taxa. They found that these peptides localized to regions that are crucial to the structural stability and function of collagen fibrils [13]. These regions contain few acidic amino acids, thus resulting in their limited solubility; furthermore, they are protected from proteolytic degradation by the close packing of collagen molecules [13].

* Corresponding author at: National Key Laboratory of Crop Genetic Improvement, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China. Fax: +86 27 87280877.
E-mail address: zhy630@mail.hzau.edu.cn (H.-Y. Zhang).

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.04.122

Nevertheless, due to the fact that amino acids are basic components of proteins, and the temperature is the main decisive factor for the survival of collagens [8,9], we speculated that, in addition to the rich content and structural stability, sufficient thermal stability of amino acids is essential for the survival of type I collagen peptides in the complex thermal history.

Due to the distinct side chains, the thermal stabilities of 20 canonical amino acids are also much different. Gly and Ala are the most stable ones [14], whereas others decline in the following order: Val ~ Leu ~ Phe > Ile > Tyr > Lys > His > Met > Thr > Ser > Trp > Glu > Asp > Arg ~ Cys [14–17]. As Gln and Asn undergo deamidation at high temperature [14], they should be less stable than Glu and Asp, respectively. Pro is rather stable, and could be compared with Phe [18].

The consistency between the stability of peptides and their amino acids is well represented in fossil-derived collagens. Fossil-derived collagen peptides are rich in Ala (~30%), Gly (~10%), and Pro (~20%, including Hyp). Hence, we used the total abundance ratio of Ala, Gly, Val, Leu, Phe and Pro (AGVLFP, the six most stable residues in our amino acid thermal stability order) as an indicator for the thermal stability of a protein. The abundance of AGVLFP in collagen peptides derived from dinosaur fossils ranged from 72% to 88%, which is much higher than that in entire collagen molecules (approximately 62%, Fig. 1). The distinct thermal stability of 20 canonical amino acids could account for both of the enrichment of Ala, Gly, and Pro and the lack of acidic residues (Asp and Glu) in dinosaur fossil-derived collagen peptides. In addition, the abundance of AGVLFP in mammoth fossil-derived type I collagen peptides (approximately 72%) is also higher than that in entire collagen molecules (approximately 62%, Fig. 2). Moreover, the abundance of AGVLFP in all mammoth fossil-derived peptides (ruling out repeatedly detected peptides and type I collagen peptides) was elevated as well, compared with that in their corresponding entire proteins (Fig. 2). Taken together, it seems that the peptides with a higher abundance of thermostable amino acids in the same protein are more likely to survive compared with the others.

These observations also bring up the question: Is the abundance of thermostable residues in type I collagens also elevated in comparison with other proteins? If so, it becomes clear why type I collagens are predominant in fossils and can be sequenced even from dinosaur fossils. To verify this, we conducted a survey of amino acid composition of all proteins in several extant species. Seven vertebrate proteomes were selected, including the proteomes of *Danio rerio*, *Xenopus tropicelis*, *Anolis carolinensis*, *Gallus gallus*, *Mus musculus*, *Loxodonta africana*, and *Homo sapiens*. Among these, *G. gallus* and *A. carolinensis* are closely related to *T. rex* and *B. canadensis* [4,5] whereas *L. africana* is closely related to the mammoth. Type I collagens are highly conserved in these organisms [4,5]. As expected, in each species, the abundance of AGVLFP in type I collagens was around 60%, ranked from the top 0.09% to

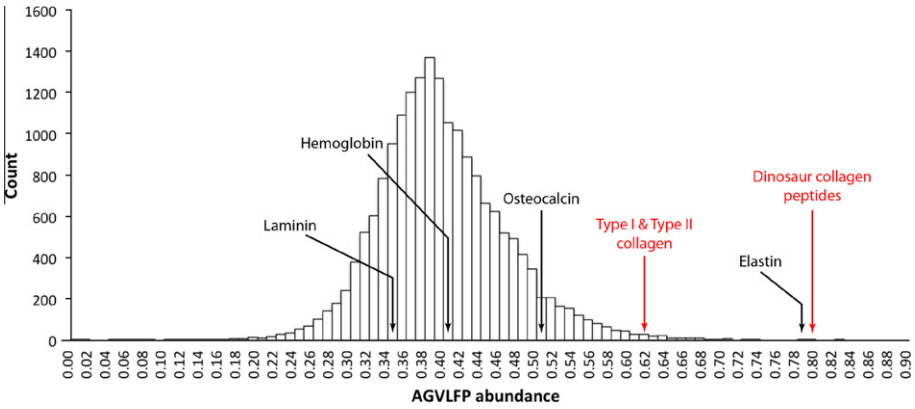

Fig. 1. Distributions of AGVLFP abundance in all proteins of *G. gallus* and the corresponding positions of dinosaur fossil-derived collagen peptides and several homologous proteins.

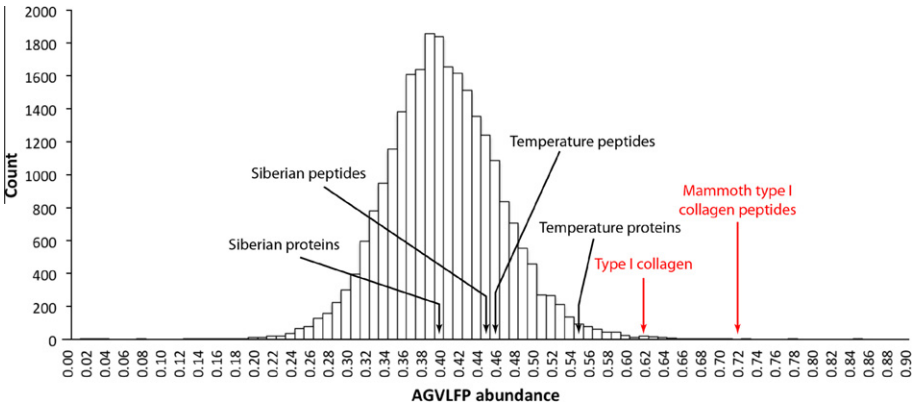

Fig. 2. Distributions of AGVLFP abundance in all proteins of *L. africana* and the corresponding positions of mammoth fossil-derived peptides. "Siberian peptides" and "Siberian proteins" represent the peptides (ruling out repeatedly detected peptides and type I collagen peptides) derived from a Siberian mammoth fossil and the corresponding entire proteins found in the proteome of *L. africana*, respectively. "Temperature peptides" and "Temperature proteins" are similar to "Siberian peptides" and "Siberian proteins," respectively, however, the peptides are derived from fossils of temperature mammoths. The "Siberian proteins" are located in the middle of the distribution, and the "Temperature proteins" have a higher abundance of AGVLFP. In addition, the abundance of AGVLFP in fossil-derived peptides in both groups is elevated.

Table 1
AGVLFP abundance for type I collagens of seven vertebrates.

Species	AGVLFP abundance[a] (%)	Rank in proteome (%)
Danio rerio	61.02; 61.70; 63.24	0.15; 0.13; 0.10
Xenopus tropicelis	62.26	0.15
Anolis carolinensis	58.70; 62.07	0.25; 0.11
Gallus gallus	64.71	0.31
Mus musculus	62.90; 63.34	0.09; 0.07
Loxodonta africana	63.23; 63.76	0.13; 0.10
Homo sapiens	63.59; 64.42	0.12; 0.09

[a] For different chains of type I collagens.

the top 0.31% in their respective proteome (Figs. 1 and 2, Table 1). Fig. 1 shows the position of dinosaur fossil-derived collagen peptides in the distribution of AGVLFP abundance in all proteins of *G. gallus*, and Fig. 2 shows the position of mammoth fossil-derived collagen peptides in the distribution of AGVLFP abundance in all proteins of *L. africana*. In the proteome of *G. gallus*, the abundance of AGVLFP in the elastin outweighed that in type I collagens (Fig. 1). The non-detection of the elastin in dinosaur collagen sequencing by mass spectrometry may be due to its rarity in the bone and the loss from the samples. Meanwhile, the abundance of AGVLFP in all other proteins, also detected using in situ immunohistochemical analyses, were much lower (Fig. 1), including osteocalcin which was also found to be abundant in the bone (1% to 2% of the total bone proteins by weight) [8]. Therefore, we speculated that the lower abundance of thermostable amino acids reduced the survival of osteocalcin and other proteins because of the deterioration of numerous residues, and only the original epitopes of these proteins were preserved whereas the molecules had been degraded. The absence of osteocalcin in the mammoth fossil-derived peptides [7] is in agreement with this hypothesis.

It is elusive why type I collagens have elevated abundance of thermostable amino acids. Due to the function of supporting tissues, large energy consumption for synthesizing and difficulties from degradation, type I collagens need to maintain integrity over a long period of time. Therefore, they are under powerful selective pressure, which constrains their sequences. An interesting finding was that, warm-water fish tended to have a similar abundance of Pro and Hyp in type I collagens to mammals, but a higher abundance compared with cold-water fish [11]. The elevated abundance of thermostable residues suggested the enhanced adaptability of type I collagens in high-temperature environment.

In summary, the present analysis provides new insights into the preservation of type I collagen, in terms of thermal stability of amino acids. Type I collagens had an extraordinarily elevated abundance of thermostable residues in comparison with other proteins in several extant species. Given that thermostable residues contributed to maintaining the integrity of type I collagens, the abundant thermostable residues in type I collagens is likely a critical factor responsible for its survival in dinosaur fossils.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 project, Grant 2010CB126100), the National Natural Science Foundation of China (Grant 21173092) and the Fundamental Research Funds for the Central Universities (Grants 2011PY142 and 2011PY027). E.C. is supported by the European Union with a Marie Curie Intra European Fellowship (Grant No. 237227).

References

[1] News Staff. Insights of the decade. Stepping away from the trees for a look at the forest. Introduction, Science 330 (2010) 1612–1613.
[2] J.R. Lloyd, C.I. Pearce, V.S. Coker, et al., Biomineralization: linking the fossil record to the production of high value functional materials, Geobiology 6 (2008) 285–297.
[3] A.R. David, M.R. Zimmerman, Cancer: an old disease, a new disease or something in between? Nat. Rev. Cancer 10 (2010) 728–733.
[4] M.H. Schweitzer, Z. Suo, R. Avci, et al., Analyses of soft tissue from *Tyrannosaurus rex* suggest the presence of protein, Science 316 (2007) 277–280.
[5] M.H. Schweitzer, W. Zheng, C.L. Organ, et al., Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. Canadensis*, Science 324 (2009) 626–631.
[6] M. Bern, B.S. Phinney, D. Goldberg, Reanalysis of *Tyrannosaurus rex* mass spectra, J. Proteome Res. 8 (2009) 4328–4332.
[7] E. Cappellini, L.J. Jensen, D. Szklarczyk, et al., Proteomic analysis of a Pleistocene mammoth femur reveals more than one hundred ancient bone proteins, J. Proteome Res. 11 (2011) 917–926.
[8] M.J. Collins, C.M. Nielsen–Marsh, J. Hiller, et al., The survival of organic matter in bone: a review, Archaeometry 44 (2002) 383–394.
[9] C.N. Trueman, D.M. Martill, The long-term survival of bone: the role of bioerosion, Archaeometry 44 (2002) 371–382.
[10] S. Perumal, O. Antipova, J.P. Orgel, Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis, Proc. Natl. Acad. Sci. 105 (2008) 2824–2829.
[11] P. Szpak, Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis, J. Archaeol. Sci. 38 (2011) 3358–3372.
[12] S.S. Raman, R. Gopalakrishnan, R.C. Wade, V. Subramanian, Structural basis for the varying propensities of different amino acids to adopt the collagen conformation, J. Phys. Chem. B 115 (2011) 2593–2607.
[13] J.D. San Antonio, M.H. Schweitzer, S.T. Jensen, et al., Dinosaur peptides suggest mechanisms of protein survival, PLoS One 6 (2011) e20381.
[14] R. Jaenicke, What ultrastable proteins teach us about protein stabilization, Biochemistry 63 (1997) 312–321.
[15] R. Jaenicke, Stability and stabilization of globular proteins in solution, J. Biotechnol. 79 (2000) 193–203.
[16] J.K. Blodgett, G.M. Loudon, K.D. Collins, Specific cleavage of peptides containing an aspartic acid (β-hydroxamic acid) residue, J. Am. Chem. Soc. 107 (1985) 4305–4313.
[17] T. Geiger, S. Clark, Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation, J. Biol. Chem. 262 (1987) 785–794.
[18] E.N. Trifonov, The triplet code from first principles, J. Biomol. Struct. Dyn. 22 (2004) 1–11.