ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 11, pp. 1227-1232. © Pleiades Publishing, Ltd., 2011.
Published in Russian in Biokhimiya, 2011, Vol. 76, No. 11, pp. 1525-1531.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM11-155, September 11, 2011.

Identification and Secondary Structure Analysis of a Keratin-Like Fibrous Protein Discovered in Ligament of the Bivalve *Siliqua radiata*

Zengqiong Huang and Gangsheng Zhang*

College of Materials Science and Engineering, Guangxi University, 100 Daxue Road, Nanning 530004, China; fax: +86-771-3270152; E-mail: huangzengqiong@163.com; zhanggsgxu@gmail.com

Received May 18, 2011
Revision received May 31, 2011

Abstract—A novel keratin-like fibrous protein K58 with molecular weight of about 58 kDa was discovered in bivalve *Siliqua radiata* ligament and identified by amino acid composition and MALDI-TOF-TOF analysis. We found that the protein is composed of cylindrical fibers (~160 nm in diameter) and contains high glycine (27.4%) and phenylalanine (10.5%) contents. Furthermore, it is homologous to keratin type II cytoskeletal 1, with repeat motifs of SGGG and SYGSGG. FTIR and secondary structure analysis indicate that K58 is composed of 46.2% β-sheet, 33.4% β-turn, 13.1% α-helix, and 4.7% disordered structure. This structure feature is closely related to the superior tensile strength, elasticity, and solvent resistance property of K58. These discoveries provide some evidence for evolution of keratin and fibrous proteins and prompt further studies of ligament fibrous proteins.

DOI: 10.1134/S0006297911110046

Key words: bivalve ligament, fibrous protein, keratin, *Siliqua radiata*, K58, secondary structure

Bivalve ligament is a rubber-like elastic structure joining the two valves dorsally and acting as a spring that causes the valves to open when adductor muscles relax. It is made up of two layers: an outer protein layer and an inner aragonite fiber layer [1, 2]. Normally, an intact ligament is composed of about 60% $CaCO_3$ and 40% protein [3]. The respective proteins have been studied for many decades.

Earlier studies were focused on amino acid composition analysis of ligament proteins. Pioneering work was carried out by Beedham, who determined the amino acid composition of *Mytilus edulis* ligament protein [4]. Later, Kelly and Rice reported a rubber-like protein, abductin, that has high glycine and methionine contents [5]. Subsequent studies investigated ligament proteins of some species [6, 7] and crystal sheath protein [8] as well as desmosine and isodesmosine in ligament [9]. In recent decades, studies were carried out mainly on the primary and secondary structure analysis of abductin and on synthesis of peptides inspired by abductin [10-12]. These studies were focused on amorphous proteins, and no fibrous proteins have been investigated. Recently, we discovered a novel spring-like fibrous protein in *Solen grandis* ligament [13] and named it solenin. It contains high Asp content and is homologous to keratin (Huang and Zhang, unpublished). Interestingly, another tensile fibrous protein (TFP) was also found in *Siliqua radiata* ligament (Fig. 1). It is composed of cylindrical fibers (~160 nm in diameter) and has never been characterized.

Here we identify and analyze the secondary structure of TFP by amino acid, MALDI-TOF-TOF, and Fourier transform infrared spectroscopy (FTIR) analysis. We found that TFP (~58 kDa) has high glycine and phenylalanine contents and contains repeat motifs of SGGG and SYGSGG. Furthermore, it is predominantly composed of β-sheet (46.2%) and turn (33.4%), with small amounts of α-helical (13.1%) and disordered (4.7%) structures.

MATERIALS AND METHODS

Sample preparation. *Siliqua radiata* specimens (Fig. 1a) were freshly collected from Beibu Gulf in southern China. We removed the soft bodies, isolated ligaments from shells, and separated the fibrous protein (Figs. 1b and 1c) from ligaments using a single-edge blade. After being

* To whom correspondence should be addressed.

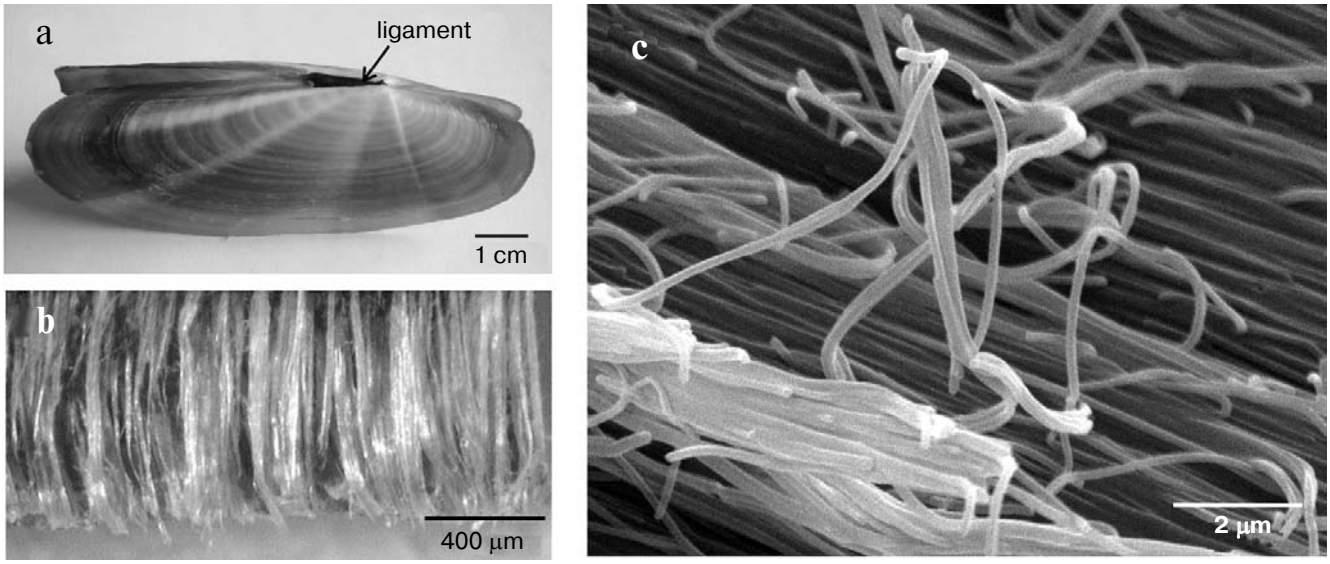

Fig. 1. Optical photos and SEM image of experiment sample: a, b) optical photos of *S. radiata* and TFP, respectively; c) SEM image of TFP.

washed with deionized water and dried at 37°C overnight, the protein was ground into powder in liquid nitrogen.

Amino acid analysis. Protein powder samples were hydrolyzed by 6 M HCl in a sealed tube under nitrogen at 110°C for 22 h. After the acid was removed, the hydrolysate was dried under vacuum, redissolved in sodium citrate buffer solution (pH 2.2), and analyzed with a Hitachi L 8800 amino acid analyzer (Hitachi, Japan).

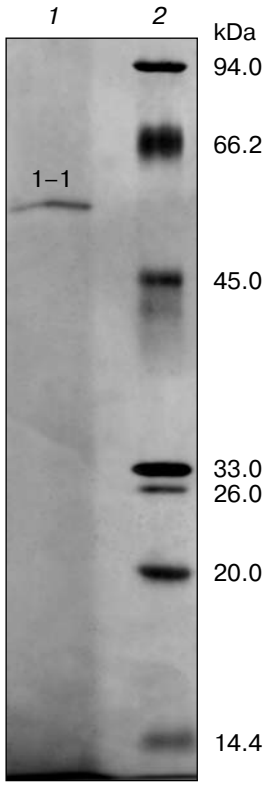

Fig. 2. SDS-PAGE pattern of TFP (*1*) and molecular weight standards (*2*).

SDS-PAGE. We dissolved the protein powder samples (15 mg) with 7 M urea solution containing 3% (v/v) 2-mercaptoethanol and 0.5 M NaOH at 70°C for 1 h. Then we centrifuged the solution for 20 min at 4°C and 12,000 rpm, and dialyzed the supernatant against deionized water in a dialysis bag of 14 kDa molecular weight cut-off (Solarbio Life Science, USA) for three days with frequent water changes. The dialysate was vacuum dried to obtain pure protein.

The protein (in Milli-Q water) and protein standards (14.4-94 kDa; TianGen Biotech, USA) were applied to SDS-PAGE on 4% stacking gel and 12% separating gel (0.1 × 13.8 × 13 cm) with a JY 600 electrophoresis system (JunYi Mechanical & Electrical Technology, China). After silver staining, we excised the gel band 1-1 (~58 kDa) in lane *1* (Fig. 2) for further analysis.

In-gel trypsin digestion. Gel band 1-1 was destained with 30 mM potassium ferricyanide containing 100 mM sodium thiosulfate, washed with 40 mM NH_4HCO_3, dehydrated with 100% acetonitrile, dried under vacuum, and rehydrated with 10 ng/μl trypsin in 40 mM NH_4HCO_3 and 10% (v/v) acetonitrile. After being incubated for 16 h at 37°C, peptides were extracted from the gel twice with 50% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid. The extract was vacuum dried.

MALDI-TOF-TOF analysis. The dried extract was redissolved in 0.1% (v/v) trifluoroacetic acid, mixed with sinapinic acid matrix, and analyzed on a mass spectrometer (model 4800; Applied Biosystems, USA). The spectrometer was operated in positive ion reflection mode in

the mass range of 800-3500 Da. The five most intense ions were automatically selected for MS-MS analysis with acceleration voltage of 20 kV.

The acquired spectra were combined and searched against mollusk and IPI_human protein database using the Mascot search engine (www.matrixscience.com). Variable modifications such as oxidation and methylation were taken into consideration, and the protein score was used for correct identification.

FTIR and secondary structure analysis. Protein powder samples were mounted in a KBr pellet and scanned using an FTIR (Nicolet 4700; Thermo Electron, USA). The spectrum was collected with 64 scans and resolution of 2 cm^{-1} in the range of 3500-400 cm^{-1}. The Amide I region (1700-1600 cm^{-1}) was curve fitted into Gaussian line shapes for secondary structure analysis by Origin 7.5 software (OriginLab, USA). Absorption band positions of individual components were used to identify α-helix, β-sheet, β-turn, and disordered structure of the protein. The percentages of different types of secondary structure were calculated by adding the sum of peak areas for each and expressing their sums as a fraction of the total Amide I band area.

RESULTS AND DISCUSSION

Amino acid composition. Table 1 gives the amino acid composition of TFP and other fibrous or elastic proteins for comparison. When compared with solenin and abductin, Gly and Phe contents are similar, but Met content, which is strikingly high in most ligaments [5-7], is only about one seventh and one twelfth of that of solenin and abductin. In addition, the difference of Asp content between TFP and solenin indicates that they are two different fibrous proteins. When compared with other proteins, Glu and Cys contents of keratin and Ala content of elastin and silk fibroin are much higher than TFP. These distinctions, along with the extremely low hydroxyproline (Hyp) content, suggest that TFP is likely a novel fibrous protein.

MALDI-TOF-TOF analysis. MALDI-TOF-TOF analysis (mass spectrum in Fig. 3) shows the expected result that no proteins of bivalve mollusk were matched when searched against the mollusk protein database. Surprisingly, keratin type II cytoskeletal 1 (KRT1) in the human protein database was matched (Table 2) with sequence coverage of 27% and protein score of 77 representing 99.8% confidence. Furthermore, the protein contains repeat motifs of SGGG and SYGSGG (Table 2) similar to the elastic repeat motifs GPGGG of byssal threads [17], which suggests that they should have similar mechanical properties. These matching motifs, with many repetitive S and G, support the amino acid analysis result, i.e. TFP has high Ser and Gly contents (Table 1). Besides, Ser content and molecular mass of TFP (~58 kDa) are similar to those of wool keratin (~60 kDa) [18]. These results indicate that TFP is a novel keratin-like fibrous protein, and it should have some relation with KRT1.

Table 1. Amino acid composition (g/100 g) of TFP and other fibrous or elastic proteins

Amino acid	TFP (*S. radiata*)	Solenin (*Solen grandis*)	Abductin (*Pecten irradians*) [5]	Keratin (wool) [14]	Elastin (bovine) [15]	Silk fibroin (*Bombyx mori*) [16]
Asp	5.43	14.15	2.62	7.88	0.53	2.53
Thr	1.24	1.15	1.44	5.57	1.4	1.14
Ser	7.69	3.46	7.54	7.9	0.9	13.84
Glu	3.77	4.45	1.95	14.93	1.94	3.48
Pro	5.16	3.75	1.54	5.53	13.78	0.92
Gly	27.35	33.44	45.23	4.01	22.24	30.96
Ala	1.07	2.3	2.94	3.79	18.2	30.74
Cys	0.78	1.69	–	13.17	0.52	0.04
Val	5.24	4.23	0.75	4.94	15.21	5.58
Met	1.21	8.84	14.93	0.58	0	0.18
Ile	2.68	1.45	1.07	3.12	3.39	1.59
Leu	0.78	2.83	0.26	7.04	7.31	1.03
Tyr	1.91	3.42	0.31	4.93	1.88	3.48
Phe	10.48	8.04	16.46	2.89	9.94	2.42
Lys	2.34	2.17	1.29	3.65	1.48	0.24
His	1.67	0.56	0.99	1.05	0	0
Arg	1.00	1.49	0.67	9.03	1.26	1.84
Hyp	0.01	0.01	–	–		–

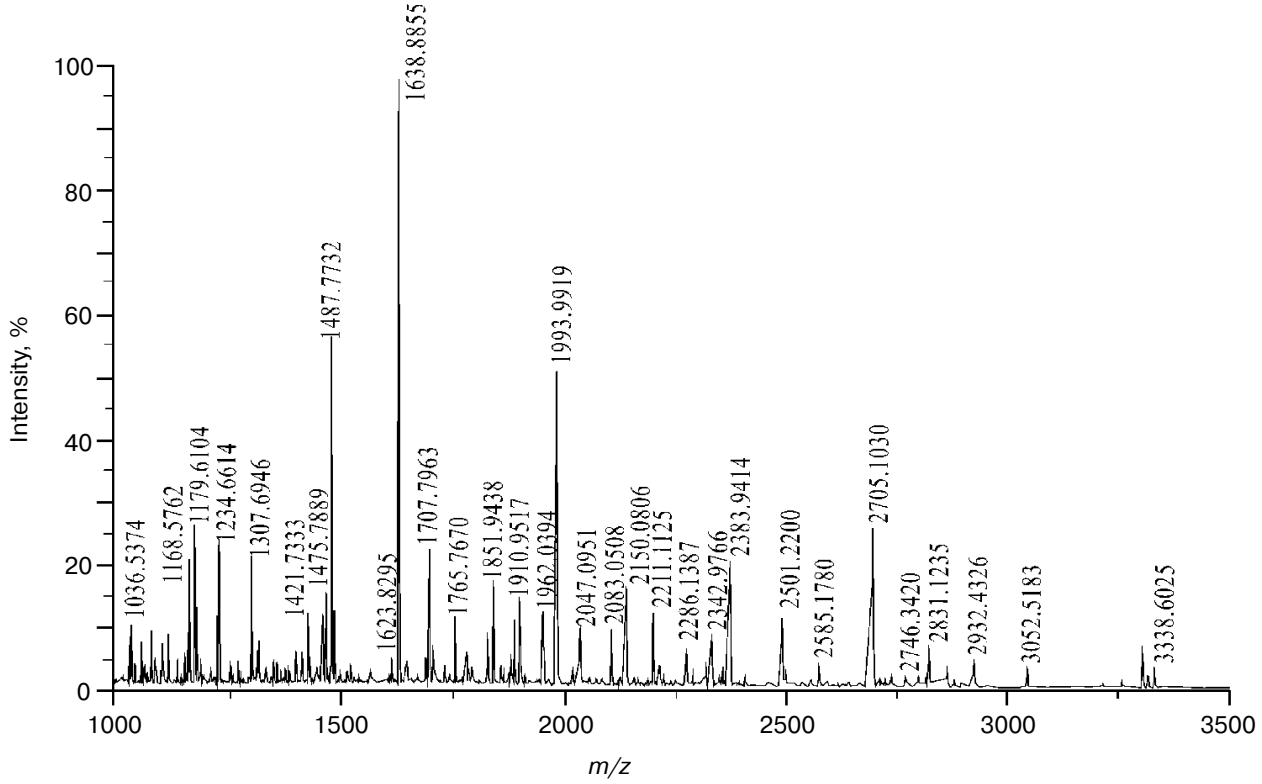

Fig. 3. MALDI-TOF mass spectrum of TFP.

KRT1 is a protein of keratin family, and type II cytokeratin is specifically expressed in the spinous and granular layers of epidermis [19]. Interestingly, the outer layer of ligament, where TFP exists, is secreted by the epithelium of the outer surface of the mantle [20]. This similarity in origin suggests that TFP is likely differentiated from cytoskeletal protein, and it should be a structural protein of the keratin family. The findings that cytoskeletal proteins exist in sea urchin and are conserved in evolution, structure, and function [21] imply that TFP is homologous to KRT1. Since it is a novel fibrous protein with MW of about 58 kDa, we named it K58. This dis-

Table 2. Search result matching keratin type II cytoskeletal 1 (KRT1) [*Homo sapiens*], IPI_human database ID: IPI00220327

M_r		Position	Peptide sequence
calculated	observed		
1033.516	1033.5172	484-492	TLLEGEESR
1141.5149	1141.5596	464-472	DYQELMNTK
1357.6985	1357.7311	444-455	LNDLEDALQQAK
1421.7383	1421.7333	277-288	RTNAENEFVTIK
1657.793	1657.8105	13-29	SGGGFSSGSAGIINYQR
1765.7349	1765.7656	46-65	FSSC*GGGGGSFGAGGGFGSR
2150.0779	2150.0798	224-240	THNLEPYFESFINNLRR
2286.1248	2286.1375	367-386	AEAESLYQSKYEELQITAGR
2932.5146	2932.4312	200-223	FLEQQNQVLQTKWELLQQVDTSTR
3312.3083	3312.1589	550-588	GSYGSGGSSYGSGGGSYGSGGGGGGGHGSYGSGSSSGGYR

Note: C*, carbamidomethylated cysteine; underlined peptides are the repetitive sequences.

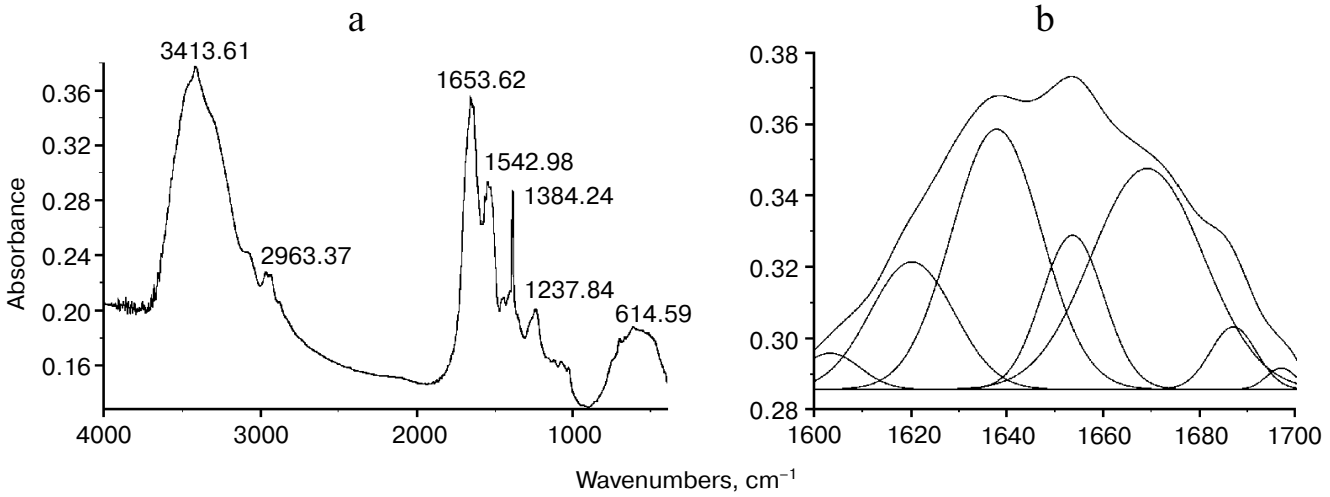

Fig. 4. FTIR spectra of K58: a) original spectrum; b) FTIR peak resolution of Amide I spectral region, 1700-1600 cm^{-1}.

covery will provide some evidence for evolution of keratin and fibrous proteins. Although such a fibrous protein does not exist in most reported ligaments, e.g. *Tellina tenuis*, *Ostrea edulis* and *Pedalion alata* [22-24], it may be ubiquitous in some families of bivalve since it was also found in ligaments of *Sinonovacula constricta* and *Solen grandis*. These two species, share common characteristics with *S. radiata*, i.e. they have two thin fragile shells and are able to burrow rapidly in the sand. It seems likely that ligament fibrous proteins are closely related to the life habit of these species. These findings will also give insight into classification of bivalves.

FTIR spectrum and secondary structure analysis. The FTIR spectrum of K58 (Fig. 4a) shows characteristic absorption bands of Amide I, II, and III at 1653, 1542, and 1237 cm^{-1}, respectively [25]. It also presents N–H stretching vibration couple with -OH peak at 3413 cm^{-1} and C–H stretching or deformation vibration at 2963, 1384, and 614 cm^{-1} [26]. This spectrum is similar to solenin [13], which suggests that they should have similar secondary structure.

For secondary structure analysis, the Amide I band of the spectrum was curve fitted into Gaussian line shapes (Fig. 4b). The peak positions of each individual component were assigned to α-helix (1650-1657 cm^{-1}), β-sheet (1612-1640 cm^{-1}), β-turn (1655-1675 cm^{-1}), and disordered (1640-1651 and 1670-1697 cm^{-1}) structure [18, 27]. Calculation results of peak area show that K58 has strikingly high β-sheet (46.2%) and turn (33.4%), with 13.1% α-helix and 4.7% disordered structure. The high β-sheet content, which is similar to wool keratin and spider silk [18, 28], we believe, is closely related to the mechanical properties of K58 and plays an important role in the ligament function.

Structure and function analysis. As mentioned, bivalve ligament functions as a spring to open the valves when adductor muscles relax. This function is mainly related to the flexible outer layer of ligament. Protein fibers of K58, existing in this layer, are subjected to tensile stress when the valves are closed. Then, the elastic energy stored in these stretched fibers causes the valves to open as adductor muscles relax. This energy enables the species to burrow in the sandy bottom by effectively opening the valves [13, 29] and is accumulated due to the high content of extensible β-sheet structure. It is accepted that β-sheet structure is closely related to the excellent elasticity and tensile strength of fibrous protein [25, 30]. For example, wool keratin, with 37.9% β-sheet structure, has a tensile strength of 0.2 GPa [18, 30]. Similarly, K58, with 46.2% β-sheet structure, should also have superior elasticity and tensile strength to effectively open the valves and withstand long term repeated tensile stress with the frequent movement of shell valves.

On the other hand, K58 has excellent chemical resistance. It can withstand exposure to concentrated hydrochloric acid, urea, and dimethyl sulfoxide at room temperature. Even 0.5 M NaOH containing 7 M urea and 3% (v/v) 2-mercaptoethanol cannot dissolve it completely at 70°C for 24 h. This property indicates poor solubility of K58, and it is due to the high contents of hydrophobic Phe, Val, and Pro, as well as the β-sheet structure. Therefore, K58 is a natural biomacromolecular material with excellent elasticity, tensile strength, and solvent resistance. It may inspire the biomimetic synthesis of elastic biomaterials.

Although this study identifies and analyzes the secondary structure of K58, further investigations are needed to obtain fundamental data of the protein. Gene cloning is in urgent need to determine the complete amino acid sequence, and tensile strength determination as well as structure prediction should be done to reveal the structure–function relationships of this unique fibrous protein.

In conclusion, K58 is a novel keratin-like fibrous protein containing SGGG and SYGSGG repeats. Secondary structure analysis reveals that it is predominantly composed of β-sheet and turn structures. This structure feature is likely to be closely associated with the superior tensile strength, elasticity, and solvent resistance of K58.

We thank the Research Center for Proteome Analysis, Key Laboratory of Proteomics, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences for MALDI-TOF-TOF determination. This work was funded by the National Natural Science Foundation of China (Grant No. 40772033).

REFERENCES

1. Trueman, E. R. (1950) *Q. J. Micro. Sci.*, **91**, 225-234.
2. Sartori, A. F., and Ball, A. D. (2009) *J. Mollusc. Stud.*, **75**, 295-304.
3. Kahler, G. A., Sass, R. L., and Fisher, F. M. (1976) *J. Comp. Physiol.*, **109**, 209-220.
4. Beedham, G. E. (1958) *Quart. J. Micro. Sci.*, **99**, 341-357.
5. Kelly, R. B., and Rice, R. V. (1967) *Science*, **155**, 208-210.
6. Kahler, G. A., Fisher, F. M., and Sass, R. L. (1976) *Biol. Bull.*, **151**, 161-181.
7. Kikuchi, Y., and Tamiya, N. (1987) *Biochem. J.*, **242**, 505-510.
8. Marsh, M., Hamilton, G., and Sass, R. (1978) *Calcif. Tiss. Res.*, **25**, 45-51.
9. Kikuchi, Y., Tsuchikura, O., Hirama, M., and Tamiya, N. (1987) *Eur. J. Biochem.*, **164**, 397-402.
10. Cao, Q., Wang, Y., and Bayley, H. (1997) *Curr. Biol.*, **7**, 677-678.
11. Bochicchio, B., Pepe, A., and Tamburro, A. M. (2005) *Chirality*, **17**, 346-372.
12. Bochicchio, B., Oronoz, F. J., Pepe, A., Blanco, M., Sandberg, L. B., and Tamburro, A. M. (2005) *Macromol. Biosci.*, **5**, 502-511.
13. Zengqiong Huang and Gangsheng Zhang (2011) *Micron*, **42**, 706-711.
14. Aluigi, A., Zoccola, M., Vineis, C., Tonin, C., Ferrero, F., and Canetti, M. (2007) *Int. J. Biol. Macromol.*, **41**, 266-273.
15. Mecham, R. P. (2008) *Methods*, **45**, 32-41.
16. Park, K. J., Jin, H. H., and Hyun, C. K. (2002) *Process. Biochem.*, **38**, 411-418.
17. Tatham, A. S., and Shewry, P. R. (2003) in *Elastomeric Proteins: Structures*, *Biomechanical Properties*, *and Biological Roles* (Shewry, P. R., Tatham, A. S., and Bailey, A., eds.) Cambridge University Press, New York, pp. 338-340.
18. Cardamone, J. M. (2010) *J. Mol. Struct.*, **969**, 97-105.
19. http://www.ncbi.nlm.nih.gov/protein/NP_006112.3.
20. Trueman, E. R. (1953) *Q. J. Micro. Sci.*, **94**, 193-202.
21. Morris, R. L., Hoffman, M. P., Obar, R. A., McCafferty, S. S., Gibbons, I. R., Leone, A. D., Cool, J., Allgood, E. L., Musante, A. M., Judkins, K. M., Rossetti, B. J., Rawson, A. P., and Burgess, D. R. (2006) *Dev. Biol.*, **300**, 219-237.
22. Trueman, E. R. (1949) *Proc. Zool. Soc. Lond.*, **114**, 717-742.
23. Trueman, E. R. (1951) *Quart. J. Micro. Sci.*, **92**, 129-140.
24. Trueman, E. R. (1954) *J. Mollusc. Stud.*, **30**, 160-166.
25. Dong, A., Huang, P., and Caughey, W. S. (1992) *Biochemistry*, **31**, 182-189.
26. Movasaghi, Z., Rehman, S., and Rehman, I. (2008) *Appl. Spectrosc. Rev.*, **43**, 134-179.
27. Pelton, J. T., and McLean, L. R. (2000) *Anal. Biochem.*, **277**, 167-176.
28. Romer, L., and Scheibel, T. (2008) in *Molecular Biology Intelligence Unit: Fibrous Proteins* (Scheibel, T., ed.) Landes Bioscience Press, Texas, pp. 121-127.
29. Sundaram, S., Deshmukh, V. D., and Raje, S. G. (2010) *J. Mar. Biol. Ass. India*, **52**, 99-101.
30. Meyers, M. A., Chen, P. Y., and Lin, A. Y. M. (2008) *Prog. Mater. Sci.*, **53**, 113.