BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(1)

Computational analysis of bovine alpha-1 collagen sequences

Ayan Saha, Md. Jibran Alam, Kutub Uddin Muhammad Ashraf & Adnan Mannan*

Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong-4331, Bangladesh; Adnan Mannan – Email: adnan_orko@yahoo.com; *Corresponding author

Received December 04, 2012; Accepted December 31, 2012; Published January 09, 2013

Abstract:
Bovine collagen alpha-1 is a naturally occurring extracellular matrix protein found in tendons and other connective tissues. It plays a vital role in cell growth, differentiation, attachment, and migration. Recent findings have established that collagen alpha-1 is involved in osteogenesis imperfecta phenotype in cattle but deep information about other members of this large family is not available so far. So with a view to finding a new edge and attempt to figure out a correlation among the well attributed Bovine alpha-1 collagen sequences are executed and analyzed. To do so, comparative analysis among the 28 members of collagen family has been carried out using Computational tools. Consequently, based on the physico-chemical, secondary structural, functional and phylogenetic classifications, we have selected collagen 12, 14 and 20 as targets for pathological conditions. These proteins belong to the FACIT family and significantly showed low glycine and proline content, high instability and aliphatic index. Moreover, FACIT family collagens contain multiple triple helical domains and being members of the FACIT family, bovine collagen 12, 14, 20 do not form fibrils by themselves but they are associated to collagen 1 associated fibrils. These collagen molecules might be crucial candidates to detect and understand the process of matrix remodeling in diseases especially in the arena of cellular compartments.

Keywords: Collagen, Extracellular matrix, computational tools.

Background:
Collagen is the most abundant family of fibrous proteins in mammals which is secreted by the connective tissue cells **[1]**. To note about its localization, it is found mostly in flesh and connective tissues in vertebrates **[2]**. Collagen structure is a triple helix with three different chains and these three alpha chains are wound around one another to form the superhelix structure which gives the long, stiff structure of collagen protein **[3]**. The amino acids in collagen are arranged in such a manner that glycine is present in every third residue **[4]**. Glycine is the smallest amino acid and thus fits perfectly in the helix and allows the alpha chains to wrap around together to form the superhelix. Collagen is rich in glycine and proline residues. So, other than glycine in every third residue, the remaining two amino acids are mostly occupied by proline. Pro-collagens are inactive precursors of collagens. During the synthesis of collagen, pro-collagens are synthesized at first. The mature active collagen molecules are formed by the action of peptidases cleaving the pro-peptides at the N and C terminals. Vitamin C acts as a cofactor in conversion of pro-collagens to collagens. Pro-collagens are cleaved only after secretion from the cells by proteolytic enzymes. Pro-collagens are fibrillar molecules which are lot more (about a thousand fold) stable than the collagen fibrils. Cleaving of pro-collagens to collagens inside the cell can lead to catastrophic consequences.

Collagen is the most abundant protein of the extracellular matrix (ECM). ECM is an intricate network of macromolecules filling the extracellular space inside the tissues. Other than collagens, ECM is rich in proteoglycans, glycoproteins and proteases **[5]**. In vertebrates, the main function of ECM is to serve as a scaffold to stabilize the physical structure of tissues.

But ECM also has more complex functions which involve cell survival, cell development, cell migration, cell-cell interaction and cell proliferation **[6]**. Evidenced and hypothetical together constitute 28 genetically distinct members of collagen protein in *Bos taurus*. In bovine, several pathological disorders are involved with imperative role of collagen. Genetic disorders in collagen synthesis include mutations in genes that encode for collagen proteins. Mutations in these genes can lead to five varieties of diseases in cattle such as Ehlers-Danlos syndrome, Osteogenesis imperfecta, Marfan syndrome, Epidermolysis bullosa (junctionalis and acanthylosis). Protein structure is the key to protein function and interaction. Protein structure analysis can provide lots of complex protein functions related disorders. Wet lab based research requires the trial and error method and cannot make a prediction before the original result. This problem can be overcome by the use of computational biology. Alteration in protein structure leads to altered protein function which in turn leads to development of diseases. So, a study that involves both dry and wet lab approaches can help to understand better about the protein function related to its structure. This type of study has been done to characterize the human matrix metalloproteinases (MMPs), in which, dry lab predictions were confirmed by experimental approaches and MMP-7 was proved as potential target in cardiac hypertrophy **[7, 8]**. Collagen acts as substrate of MMPs and is involved in many pathological conditions. A derivative of collagen, gelatin also shows such kind of relation to diseases. Analysis of collagen is thus essential to understand the process of matrix remodeling in diseases. In this specific study, analysis of bovine alpha-1 collagen sequences is done by using computational tools. Alpha-1 is present in all 28 bovine collagen. So, bovine collagen alpha-1 chain was selected for further study of collagen sequences. In our study, secondary structural, physico-chemical, phylogenetic and functional analysis of bovine alpha-1 collagen sequences were done. The target of this research is to give an insight about the nature of collagen proteins and characterize this protein family. Proposal about the potential members involved in disease conditions is also an intention of this study by examining the collagen protein family and finding any abnormal characteristics in the protein molecules. Thus further studies on collagen protein family would be facilitated by this research.

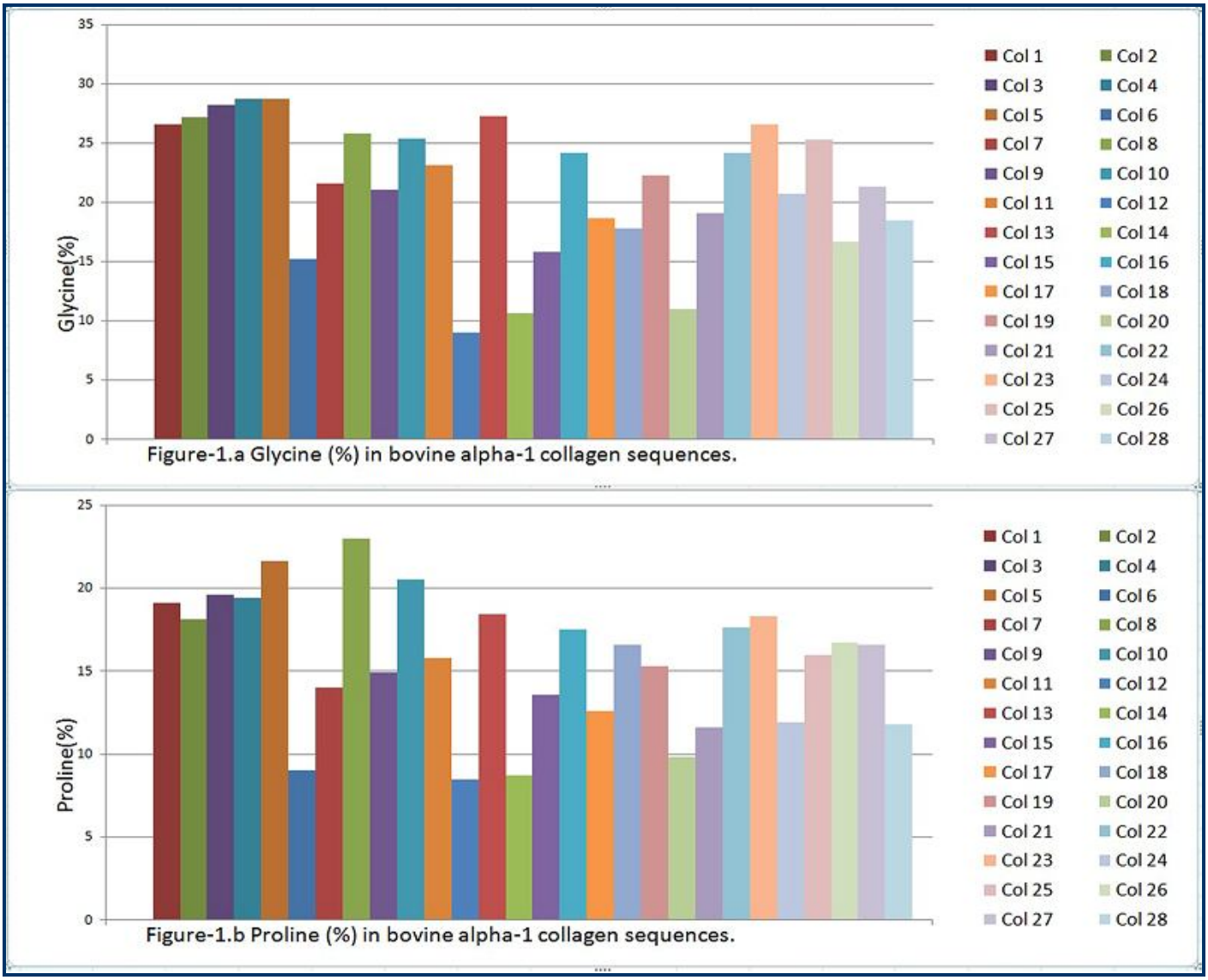

Figure 1: Glycine and Proline content (in %) of bovine alpha-1 collagens**.**

Methodology:

Protein sequence retrieval

For all 28 members of bovine alpha-1 collagen family, protein sequences were derived from protein database of NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/protein/) in FASTA format. Various other computational biology tools were used for further analysis of these sequences.

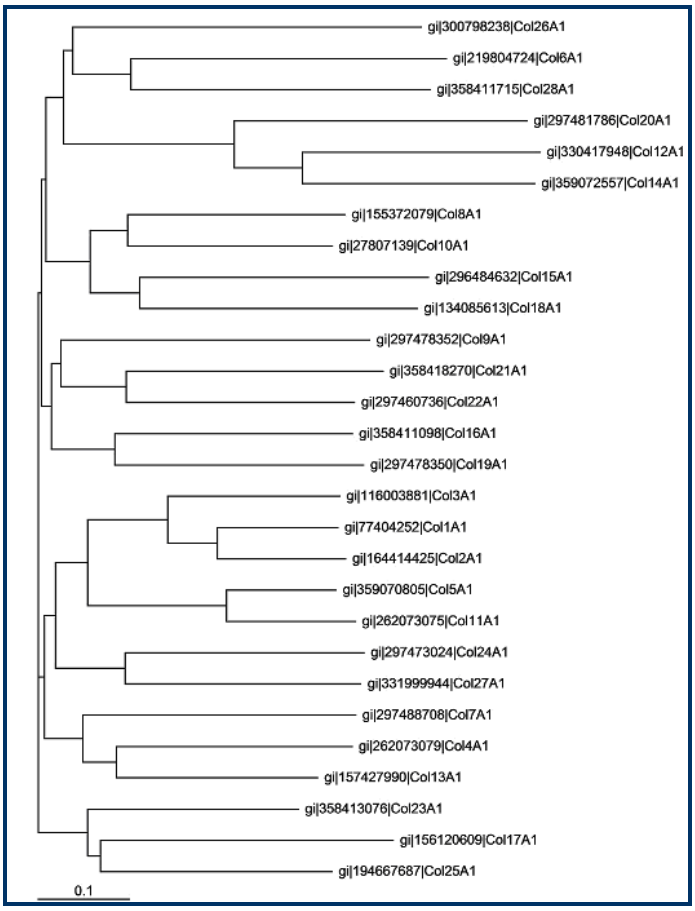

Figure 2: Phylogenetic tree of bovine alpha 1 collagen sequences by using Neighbor-Joining Method.

Analysis of Physico-chemical properties

The ProtParam tool (http://web.expasy.org/protparam/) of ExPASy was used to compute amino acid composition (%), molecular weight, theoretical isoelectric point (pI), number of positively and negatively charged residues, extinction coefficient, instability and aliphatic index, Grand Average of Hydropathy (GRAVY). Another ExPASy tool known as ProtScale (http://web.expasy.org/protscale/) was used to compute the number of codons, bulkiness, polarity, refractivity, recognition factors, hydrophobicity, transmembrane tendency, percent buried residues, percent accessible residues, average area buried and average flexibility **[9]**.

Analysis of Secondary structural properties

Secondary structural properties of the protein including alpha helix, 3_{10} helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous states and other states were computed by the use of SOPMA (Self Optimized Prediction Method with Alignment, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) tool of NPS (Network Protein Sequence Analysis) **[10]**.

Analysis of functional properties

For functional analysis, the motifs of the alpha-1 protein sequences were identified by using the Motif Scan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan) tool **[11]**. The input data type was in FASTA format and motifs were scanned against Prosite patterns.

Phylogenetic analysis

Phylogenetic analysis of bovine alpha-1 collagen sequences was done by two softwares, ClustalX and TreeView. All the sequences were aligned by using the clustalx version 2.1. Then phylogenetic tree was generated by using NJ method. The output of phylogenetic tree in Phylip format was then viewed by TreeView.

Discussion:

For all the collagen, three criteria were analyzed- the biological processes they are involved in, cellular components they are part of and their molecular function **Table 1 (see supplementary material)** Collagen 1 and 2 seem to be involved in a huge number of biological processes. No data involving the three selected criteria were found for 8 collagen members. It signifies that these collagen members are yet uncharacterized for their biological process, cellular component and molecular function. From ProtParam result, it was observed that for all the residues, the percentage of glycine and proline was higher than the other residues. Except collagen 12, 14 and 20 glycine content was higher than 12% **(Figure 1)**. High glycine content is a necessity for collagens to maintain their triple helical structure since larger amino acids cause steric hindrance **[12]**. Proline content for all members except collagen 12, 14 and 20 was observed more than 10% **(Figure 1)**. Proline residues of collagen are necessary to stabilize the helix and disrupt the structure of secondary structural elements **[13]**. So, for collagen to be a protein molecule and carry out processes like cell migration and cell adhesion, proline concentration is important. The total number of positively (Arg + Lys) and negatively (Asp + Glu) charged residues of the collagen members were observed **Table 2 (see supplementary material)**. For 14 members, the total number of positively charged residues exceeded the total number of negatively charged residues; they showed their isoelectric point in alkaline range. For the remaining members, the isoelectric point was within acidic range. Extinction coefficient for collagen 7, 11, 12, 14, 17, 18 and 20 was observed higher than remaining members. Higher extinction coefficient means higher concentration of lysine, tryptophan and tyrosine. This observation is important for protein-protein interaction studies. Whether a protein is stable or not can be described by its instability index. Instability index for collagen 14, 17, 18 and 20 is higher than 40 and thus describing these proteins as unstable **[14]**. High aliphatic index was observed for collagen 12, 14 and 20. Higher aliphatic index indicates higher concentration of alanine, valine, isoleucine and leucine occupying the relative volume of a protein **[14]**. Again, higher aliphatic index provides higher thermostability. For bovine alpha-1 collagen, aliphatic index ranges from 35.18 to 84.29. It is a wide range and suggests that most of the collagens may be stable. For collagen 14 and 20, the instability index and aliphatic index results contradict greatly. According to instability index,

these proteins are unstable but their aliphatic index is high enough to say that they are stable.

Grand Average of Hydropathy (GRAVY) was computed for all the members. A broad range of GRAVY value was observed from -0.955 to -0.223 for bovine alpha-1 collagen. By ExPASy's ProtScale tool hydrophobicity was measured and it ranges from -0.3555 for collagen 20 (most hydrophilic) to 0.1275 for collagen 23 (most hydrophobic). Average flexibility ranges between 0.4365 and 0.459; a short range which indicates high glycine and proline content in the proteins **Table 3 (see supplementary material)**. SOPMA analysis was done for all bovine alpha-1 collagen members and it showed a high value for Random coil in all the members **Table 4 (see supplementary material)**. The values for alpha helix were found higher than extended strands in 13 collagens. High value for random coil bears important significance in the study of protein tertiary structure and related functions. Collagen 1, 2, 3 motifs were described as VWFC domain signature and the other two were described as pancreatic trypsin inhibitor family signature. As VWFC domain is involved in oligomerization, so it could be related to the assembly of collagen into a triple helical structure. Furthermore, Collagen 7 and 28 were showed to have pancreatic trypsin inhibitor (Kunitz) family signature which manifest strong matches in the motifs **Table 5 (see supplementary material)**. Phylogenetic tree was constructed with distance based Neighbor-Joining method. A number of clusters were found including 6 and 28, 12 and 14, 21 and 22, 1 and 2, 4 and 13, 17 and 25 lying in close proximity to 26, 20, 9, 3, 7 and 23 respectively **(Figure 2)**. Proteins in close evolutionary relationship may be analyzed together for their involvement in similar biological processes. Collagen 1 has already been reported as a key player in cattle osteogenesis imperfecta **[15]**. The FACIT (Fibril Associated Collagens with Interrupted Triple Helices, Collagen 9, 12, 14 and 20) get associated with collagen 1 and then form fibrillar structure. In human, collagen 9 has already been reported as responsible for skeletal disorders **[16]**. Collagen 12, 14 and 20 might be potent target in pathological conditions. Based on their similarities and abnormalities in structural properties, these protein molecules might be accounted for investigation for their involvement in pathological conditions.

Conclusion:

In this research, we tried to disclose the hidden information about bovine alpha-1 collagen by analyzing their structural features e.g. amino acid content, physico-chemical properties, secondary structural features and phylogenetic classification. Various computational tools were used to ease up the process of finding. Change in protein structure can cause impairment of protein function and develop many pathological conditions. Disease conditions interfere with the normal biological processes in animals. Apart from these, based on comparative characterization and analyzing the evolutionary relationship it can be hypothesized that collagen 12, 14 may be potential target in pathological conditions and they show a close resemblance with collagen 20 in phylogenetic tree for which cellular and molecular function still not revealed. Hence it can be assumed that collagen 12 and 20 also can interact with the fibril surface and regulate fibrillogenesis which is a unique feature of collagen 14. Moreover, all of these collagens belong to FACIT collagen family and share similar properties and abnormal behaviors; e.g. they have very low percentage of glycine content, high instability index and high aliphatic index. To sum up, this experiment will provide an insight for the biologists working with ECM proteins in order to prosecuting research on collagen to find out different cell mediated injuries and so on. Findings of this study need further studies and validation by experimental research.

References:

[1] Di Lullo *et al. J Biol Chem*. 2002 **277**: 4223 [PMID: 11704682]
[2] Müller *et al. Integrated Computational Biology*. 2003 **43**: 3
[3] Szpak *et al. Journal of Archaeological Science*. 2011 **38**: 3358
[4] Kadler *et al. Biochem J*. 1996 **316**: 1 [PMID: 8645190]
[5] Järveläinen *et al. Pharmacol Rev*. 2009 **61**: 198 [PMID: 19549927]
[6] Bowers *et al. J Mol Cell Cardiol*. 2010 **48**: 474 [PMID: 19729019]
[7] Jaiswal A *et al. Bioinformation*. 2011 **6:** 23 [PMID: 21464841]
[8] Chhabra *et al. J Comput Intell Bioinformatics*. 2011 **4**: 1
[9] Gasteiger *et al. Nucleic Acids Res.* 2003 **31**: 3784 [PMID: 12824418]
[10] Geourjon C *et al. Comput Appl Biosci*. 1995 **11:** 681 [PMID: 8808585]
[11] Pagni M *et al. Nucleic Acids Res*. 2007 **35**: W433 [PMID: 17545200]
[12] van der Rest M & Garrone R, *FASEB J*. 1991 **5**: 2814 [PMID: 1916105]
[13] Ramachandran GN *et al. Biochim Biophys Acta*. 1973 **332**: 166 [PMID: 4744330]
[14] http://web.expasy.org/docs/expasy_tools05.pdf
[15] Agerholm *et al. Zentralbl Veterinarmed A*. 1994 **41**: 128 [PMID: 8091888]
[16] Czarny-Ratajczak *et al. Am J Hum Genet*. 2001 **69**: 969 [PMID: 11565064]

Edited by P Kangueane

Citation: Saha *et al.* Bioinformation 9(1): 042-048 (2013)

BIOINFORMATION open access

Supplementary material:

Table 1: Functional properties of bovine alpha 1 collagens.

NCBI-GI	Collagen	Biological process	Cellular component	Molecular function
77404252	Col 1	Blood vessel development, bone trabecula formation, cartilage development , cellular response , collagen fibril organization, positive regulation of canonical Wnt receptor signaling pathway, cell migration, skin morphogenesis etc.	Collagen type I, cytoplasm, extracelluler space etc	Extracellular matrix structural constituent
164414425	Col 2	Cartilage condensation, cartilage development involved in endochondral bone morphogenesis, collagen fibril organization, embryonic skeletal joint morphogenesis, proteoglycan metabolic process, regulation of gene expression, tissue homeostasis etc.	Basement membrane, collagen type II, cytoplasm, extracellular space etc.	Extracellular matrix structural constituent
116003881	Col 3		Collagen	Extracellular matrix structural constituent
262073079	Col 4		Collagen	Extracellular matrix structural constituent
359070805	Col 5		Collagen	Extracellular matrix structural constituent
219804724	Col 6	Cellular response to amino acid stimulus, protein heterotrimerization etc.	Extracellular matrix, extracellular region, protein complex sarcolemma etc.	
297488708	Col 7		Basement membrane	Serine-type endopeptidase inhibitor activity
155372079	Col 8	Camera-type eye morphogenesis epithelial cell proliferation, positive regulation of cell-substrate adhesion etc.	Collagen	
297478352	Col 9	Tissue homeostasis, growth plate cartilage development, cartilage development etc.	Collagen type IX	
27807139	Col 10		Extracellular region, proteinaceous extracellular matrix collagen etc	
262073075	Col 11			Extracellular matrix structural constituent
330417948	Co 12		Extracellular matrix, extracellular space etc	
157427990	Col 13	Cell-cell adhesion, cell-matrix adhesion, endochondral ossification, morphogenesis of a branching structure etc	Cell-cell junction	Heparin binding
359072557	Col 14			Extracelluler space, extracelluler matrix etc.
296484632	Col 15	Cell adhesion	Collagen, extracellular space etc.	Structural molecule activity
358411098	Col 16	Cellular response to amino acid stimulus		
156120609	Col 17	Hemidesmosome assembly	Basement membrane collagen, hemidesmosome integral to membrane etc.	
134085613	Col 18	Cell adhesion	Collagen	Structural molecule activity
297478350	Col 19			
297481786	Col 20			
358418270	Col 21			
297460736	Col 22			
358413076	Col 23			
297473024	Col 24			
194667687	Col 25		Extracellular space, integral to plasma membrane etc.	Heparin binding
300798238	Col 26	Positive regulation of cell-substrate adhesion	Golgi apparatus, endoplasmic reticulum, proteinaceous extracellular matrix etc	
331999944	Col 27		Fibrillar collagen	Extracellular matrix structural constituent
358411715	Col 28		Basement membrane	Serine-type endopeptidase inhibitor activity

***Col- Collagen.**

Table 2: Physico-chemical parameters of bovine alpha-1 collagens.

Collagens	No. of amino acids	Molecular weight	pI	'-' charged residues	'+' charged residues	Extinction Coefficient	Instability index	Aliphatic index	GRAVY
Col 1	1463	138938.4	5.6	140	127	57965	31.64	38.09	-0.799
Col 2	1487	141828.4	6.82	141	140	54525	25.52	38.98	-0.810
Col 3	1466	138438.7	6.06	128	119	63715	29.93	35.18	-0.822
Col 4	1669	160430.1	8.56	128	138	56600	31.78	45.35	-0.650
Col 5	1087	104339.8	4.94	122	93	20985	31.75	37.38	-0.955
Col 6	1027	108670.7	5.24	142	127	63480	27.83	68.09	-0.544
Col 7	2932	293345.4	6.25	329	317	182755	32.10	61.97	-0.624

Col 8	745	73253.7	9.59	38	60	38405	34.45	60.72	-0.456
Col 9	940	94675.6	9.23	85	100	57700	34.44	61.05	-0.578
Col 10	674	65546.1	9.59	34	51	36455	27.93	54.42	-0.550
Col 11	1817	182416.4	5.66	210	191	100785	30.79	45.50	-0.856
Col 12	3065	332999.5	5.38	362	309	334620	34.04	76.42	-0.419
Col 13	686	66554.8	8.44	73	77	15970	30.13	48.57	-0.836
Col 14	1797	193403.9	5.10	214	162	182200	40.31	77.65	-0.335
Col 15	1386	141359.9	4.82	158	97	79340	39.74	68.20	-0.366
Col 16	1613	159151.7	7.45	152	153	61360	38.01	48.64	-0.713
Col 17	1473	149153.8	8.82	119	128	121380	44.34	55.63	-0.581
Col 18	1514	173951.2	5.33	160	124	135830	41.05	56.66	-0.606
Col 19	1195	121410.5	9.04	121	139	60360	35.92	52.58	-0.784
Col 20	1342	142754.7	8.92	120	133	127615	47.94	84.29	-0.223
Col 21	957	9916631.5	8.21	98	102	58060	31.54	67.55	-0.504
Col 22	1605	159096.5	6.81	173	170	54735	34.11	50.75	-0.768
Col 23	361	35165.5	5.82	41	38	7115	29.01	52.22	-0.827
Col 24	1467	151102.3	7.16	148	147	71335	29.07	63.96	-0.650
Col 25	673	66495.9	8.77	74	81	11960	26.48	52.07	-0.822
Col 26	438	44586.3	6.71	38	37	38680	46.52	61.78	-0.524
Col 27	1841	185300.7	9.80	140	197	74215	38.72	55.18	-0.642
Col 28	1147	119374.6	8.13	132	136	53705	28.57	62.98	-0.662

***Col- Collagen.**

Table 3: Physico-chemical properties of bovine alpha-1 collagens.

Collagen	No. of Codons	Bulkiness	Polarity	Refractivity	Recognition Factors	Hydro-phobicity	Trans-membrane Tendency	% buried residues	% accessible residues	Average Area Buried	Average Flexibility
Col 1	3.889	14.3645	8.7385	13.8225	87.5	-0.1545	-0.609	6.467	5.6245	118.0225	0.4505
Col 2	3.5	14.6425	8.411	14.417	89.722	-0.131	-0.4545	7.006	6.0775	116.2055	0.452
Col 3	3.889	13.5935	8.55	13.884	88.222	-0.092	-0.4285	7.178	5.65	117.439	0.4435
Col 4	3.611	13.0435	8.2775	14.739	89.2225	-0.1725	-0.3765	6.522	6.111	119.278	0.4375
Col 5	3.3885	14.4175	8.3335	13.7115	87.889	-0.0395	-0.487	6.128	5.9165	118.3945	0.455
Col 6	3.5555	13.9375	8.728	16.544	88.111	-0.069	-0.6105	6.089	5.672	116.5945	0.457
Col 7	3.7775	13.844	8.683	14.118	90.5555	-0.084	-0.6075	6.6945	5.822	118.2885	0.4475
Col 8	3.722	13.946	8.2725	15.9515	87.167	-0.0755	-0.453	6.5835	5.789	122.661	0.447
Col 9	3.834	13.454	8.35	15.108	88.167	-0.085	-0.314	6.306	5.383	122.934	0.449
Col 10	3.611	13.962	8.2165	14.807	87.611	-0.2	-0.379	5.783	5.561	121.122	0.453
Col 11	3.445	13.735	8.867	16.398	89.3335	-0.051	-0.694	6.372	6.633	122.972	0.45
Col 12	3.7775	14.686	8.45	15.226	90	-0.117	-0.538	7.0445	5.6725	121.2225	0.4515
Col 13	3.833	13.3785	8.4225	15.171	88.0555	0.0235	-0.5005	6.5945	6.111	118.511	0.4455
Col 14	3.4445	14.748	8.389	16.036	89.389	-0.146	-0.7235	5.561	5.761	120.728	0.445
Col 15	4.0	14.098	8.5	14.9115	88.1665	-0.094	-0.5655	6.8555	5.922	120.828	0.448
Col 16	3.5	14.113	8.489	15.22	88.0555	-0.158	-0.458	6.039	5.6555	120.1835	0.4435
Col 17	3.889	12.439	8.1945	15.4135	90.5	-0.3005	-0.342	7.039	5.9	119.3775	0.453
Col 18	3.7225	13.1815	9.344	13.5535	89.2775	-0.0025	-1.111	6.4055	5.689	115.05	0.4505
Col 19	3.4445	14.44	8.939	15.7285	87.9445	-0.003	-0.96	5.9165	5.7055	124.1665	0.443
Col 20	3.9445	14.202	8.5275	14.936	90.6665	-0.3555	-0.6615	7.0555	5.8835	120.8335	0.446
Col 21	3.5555	14.837	8.233	15.478	87.389	-0.267	-0.31	6.017	5.633	124.35	0.4365
Col 22	3.611	13.877	8.2615	15.5015	88.278	-0.2405	-0.32	6.3225	5.9445	124.6835	0.445
Col 23	3.6115	13.5165	8.878	12.1955	86.722	0.1275	-0.8055	6.7445	6.05	108.3225	0.459
Col 24	3.5	14.685	9.022	16.653	89.5555	-0.281	-1.0535	6.0445	5.673	121.223	0.452
Col 25	3.8335	13.3225	8.6335	12.9695	87.5555	0.0245	-0.617	7.083	5.9055	113.683	0.456
Col 26	3.2225	13.3085	8.417	15.4075	86.222	-0.071	-0.6255	6.489	5.1835	116.1885	0.4455
Col 27	3.4445	14.224	8.5275	14.413	88.3885	-0.096	-0.356	6.5555	5.9665	120.6055	0.445
Col 28	3.5	14.9415	8.3055	15.9065	88.611	-0.2785	-0.5095	6.278	5.8335	124.289	0.448

***Col- Collagen.**

Table 4: Secondary structural features of bovine alpha-1 collagens (in %).

Collagen	α Helix	3_{10} Helix	Pi Helix	β Bridge	Extended Strand	β Turn	Bend Region	Random Coil	Ambiguous States	Other States
Col 1	4.92	0	0	0	7.86	3.83	0	83.39	0	0
Col 2	7.06	0	0	0	7.20	4.64	0	81.10	0	0
Col 3	3.62	0	0	0	8.19	4.98	0	83.22	0	0
Col 4	4.43	0	0	0	7.73	5.75	0	82.09	0	0
Col 5	5.43	0	0	0	5.89	3.86	0	84.82	0	0
Col 6	27.17	0	0	0	13.73	5.36	0	53.75	0	0
Col 7	9.82	0	0	0	15.86	6.86	0	67.46	0	0
Col 8	3.76	0	0	0	10.87	8.59	0	76.78	0	0
Col 9	15.32	0	0	0	11.49	8.40	0	64.79	0	0
Col 10	4.15	0	0	0	11.57	5.93	0	78.34	0	0
Col 11	7.43	0	0	0	11.83	5.78	0	74.96	0	0
Col 12	12.89	0	0	0	25.77	5.22	0	56.12	0	0
Col 13	11.66	0	0	0	4.81	4.23	0	79.30	0	0
Col 14	16.14	0	0	0	23.32	4.62	0	55.93	0	0
Col 15	18.33	0	0	0	16.52	6.42	0	58.73	0	0
Col 16	9.49	0	0	0	9.49	6.32	0	74.71	0	0
Col 17	19.28	0	0	0	9.64	5.50	0	65.58	0	0
Col 18	13.74	0	0	0	12.48	6.27	0	67.50	0	0
Col 19	14.31	0	0	0	9.62	5.86	0	70.21	0	0

Col 20	16.17	0	0	0	23.55	3.50	0	56.78	0	0
Col 21	15.15	0	0	0	10.97	1.78	0	72.10	0	0
Col 22	11.84	0	0	0	10.28	5.92	0	71.96	0	0
Col 23	8.69	0	0	0	1.39	1.39	0	88.64	0	0
Col 24	6.95	0	0	0	11.32	3.41	0	78.32	0	0
Col 25	11.74	0	0	0	3.57	2.38	0	82.32	0	0
Col 26	25.11	0	0	0	9.82	5.02	0	60.05	0	0
Col 27	10.05	0	0	0	11.03	6.25	0	72.68	0	0
Col 28	19.09	0	0	0	13.16	5.32	0	62.42	0	0

***Col- Collagen.**

Table 5: Motifs in bovine alpha-1 collagens.

Collagen	Motif found	Motif ID	Description	Start	End	Match Status Significance
Col 1	VWFC_1	PS01208	VWFC domain signature	58	95	Strong match; not a false positive
Col 2	VWFC_1	PS01208	VWFC domain signature	58	89	Strong match; not a false positive
Col 3	VWFC_1	PS01208	VWFC domain signature	50	88	Strong match; not a false positive
Col 7	BPTI_KUNITZ_1	PS00280	Pancreatic trypsin inhibitor (Kunitz) family signature	2895	2913	Strong match; not a false positive
Col 28	BPTI_KUNITZ_1	PS00280	Pancreatic trypsin inhibitor (Kunitz) family signature	1122	1140	Strong match; not a false positive

***Col- Collagen.**