Identification and Computational Analysis of Chicken Alpha-1 Collagen Sequences

Z. Islam, M.M. Islam, S. Saha, C.A. Jahangir, B. Basak, M.N. Islam, M.S. Islam, S. Paul, M. Khalekuzzaman*

Abstract— Collagen is a protein that is found in cartilage, bone and other tissues in animals and humans. People utilize collagen from chicken for medicine. It is used to treat joint pain associated with many types of arthritis and surgery, with back pain, neck pain, and pain subsequent injury. So far, 26 genetically distinct collagen types have been described. However, there is not sufficient information to know if chicken collagen is safe or what the side effects might be. Other collagen products, such as bovine collagen and gelatin, have caused allergic reactions whereas chicken collagen has low amount of allergic reactions. It is recently found that collagen alpha-1 is involved in osteogenesis imperfecta phenotype in cattle, but there is no such type of information in case of chicken. So, attempting to figure out information about their structural features chicken alpha-1 collagen sequences are carried out and analyzed. This investigation reports a comparative analysis and characterization of the alpha-1 chain of chicken collagen sequences using different computational tools. Amino acid composition, physico-chemical, secondary structural, functional and phylogenetic classification were done, based on which, Alpha-1 (XIV) collagen, Collagen alpha-1(XIV) chain precursor and Collagen type XII alpha-1, which belong to the Fibril Associated Collagens with Interrupted Triple helices (FACIT) collagen family, can be recognized as potential players in diseased conditions, because of certain unusual properties such as very high aliphatic index, low percentage of Glycine and Proline residues and their proximity in evolutionary history. These collagen molecules might be play crucial role in cosmetics, foods on top of in medical industry.

Index Terms— Chicken, collagen, computational tools, phylogenetic, physico-chemical

1. INTRODUCTION

Collagen is a naturally-occurring protein which found in animal connective tissues, accounting approximately thirty percent of animal protein. However, the most abundant proteins in the extracellular matrix are members of the collagen family. So far, 19 collagen types have been identified in mammals. In all these types a major component of the protein is a triple-helical structure with three distinct polypeptide chains, generally known as the alpha chains. At least 30 genes are required to code for the constituent chains of these 19 types [1, 2]. Each polypeptide chain has a repeating Gly-X-Y triplet in which the amino acids in collagen are organized in such a way that glycine is present in every third residue [3] responsible for the stability of the helical structure due to its property of being the smallest amino acid and the X and Y positions are frequently occupied by proline and 4-hydroxyproline, respectively. Therefore, collagen is well known as glycine and proline rich entity. Inactive precursors of collagen are known as pro-collagens. Pro-collagens are synthesized at first during the synthesis of collagen. The mature active collagen molecules which are formed by peptidases action cleaving pro-peptides present at the N and C terminal. As a cofactor Vitamin C play an important role in conversion of pro-collagens to collagens. However, Pro-collagens are cleaved only after secretion from the cells by proteolytic enzymes. The diverse collagen family grouped into fibril-forming collagens, fibril-associated collagens (FACIT), network- forming collagens, anchoring fibrils, transmembrane collagens, basement membrane collagens and others with unique functions [4]**.** Collagen has been used owing to its exclusive character in assorted field of industry such as leather and films, beauty and cosmetics, biomedical, pharmaceuticals materials and foods [5]. Most commercial collagen has been extracted from mammary animals especially cattle, pig skin. Although, A researcher reported that a low amount of allergic gelatin could be produced from chicken cartilage by acid processing [6].

Consumers had recently refused some goods and cosmetics that were prepared from beef collagen or gelatin by reason of bovine several pathological disorders are involved. Genetic disorders in collagen synthesis contain mutations in genes that encode for collagen proteins. Mutations in these genes can instigate varieties of diseases in cattle for instance, Ehlers-Danlos syndrome, Osteogenesis imperfecta, Marfan syndrome, Epidermolysis bullosa (junctionalis and acanthylosis)

- *Z. Islam, M.M. Islam, S. Saha, C.A. Jahangir are students of the Dept. of Genetic Engineering and Biotechnology, Faculty of Life and Earth Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh*
- *B. Basak is a student of the Dept. of Biochemistry and Molecular Biology, Faculty of Science, University of Rajshahi, Rajshahi-6205, Bangladesh*
- *M.N. Islam, M.S. Islam, S. Paul are students of the Dept. of Pharmacy, Faculty of Science, University of Rajshahi, Rajshahi-6205, Bangladesh*
- *M Khalekuzzaman is a Professor, Dept. of Genetic Engineering and Biotechnology, Faculty of Life and Earth Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh & *Corresponding author, Email: kzaman63@gmail.com, kzaman63@ru.ac.bd*

[7]. Therefore, it is required to search for an alternative source of collagen from the other animal species rather than from cattle and chicken bi-products such as feet and skin may be a best source.

Chicken is one of the most common and widespread domestic animal which is preferred to human as a source of food, consuming both their meat and eggs. It comprises not only high-quality protein, but also important vitamins and minerals. Chickens in extensive and semi-intensive poultry production systems report for above 75% of all poultry in the South. The objective of this research was to analyze chicken collagen properties such as amino acid composition, physico-chemical, secondary structural features, phylogenetic classification and function so that it will be helpful in future to characterize alpha 1 collagen from chicken and utilize cosmetic, food and in medical application.

2. MATERIALS AND METHODS

2.1. Sequence retrieval of chicken collagen alpha-1 proteins

The protein sequences of all 8 members of chicken alpha-1 collagen family reported so far were retrieved from the national center for biotechnology information (NCBI) source in FASTA format with the following accession numbers AAF28099.1, P02457.3, CAA46238.1, AAO23053.1, NP_990665.1, AAC83579.1, 0409238A, and 0502173A.

2.2. Physico-chemical properties analysis

The protein sequence in FASTA format [8, 9] as the input data type was used to compute the 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) in Protparam tool (http://web.expasy.org/protparam/) which is available on Expasy server. In addition, other physico-chemical features including number of codons, bulkiness, polarity, refractivity, recognition factors, hydrophobicity, transmembrane tendency, percent buried residues, percent accessible residues, average area buried and average flexibility and relative mutability were evaluated using a different ExPASy tool which is well-known as ProtScale (http://web.expasy.org/protscale/) [10].

2.3. Secondary structure analysis

To analyze the secondary structural features of amino acid sequences SOPMA tool (Self-Optimized Prediction Method with Alignment) of NPS (Network Protein Sequence Analysis, http://npsapbil.ibcp.fr/cgibin/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) server was used to illustrate alpha helix, 3_{10} helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous states and other states [7].

2.4. Functional properties analysis

The analysis of the selected alpha-1 protein sequences of the collagen family were done with the help of Motif scan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan) [8]. The input data type was in FASTA format and scanned against 'PROSITE Patterns' which is a selected protein profile database out of the eight available.

2.5. Phylogenetic analysis

The chicken alpha-1 collagen protein sequences were aligned using multiple sequence alignment tool ClustalW2 where the protein sequences in FASTA format as the input data type (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al. 2007**).** Phylogenetic tree was constructed using Neighbor Joining (NJ) method which helps to set up evolutionary relationships.

3. RESULTS AND DISCUSSION

Using ExPASY's ProtParam tool the computation of amino acid composition of each chicken alpha-1 collagen sequence signified very high percentages of glycine and proline as compared to other amino acids. Glycine and proline content in all selected collagens was more than higher than the other residues. Except collagen alpha-1(XIV) chain precursor and collagen type XII alpha-1 glycine content was higher than 12% , with a value of 10.6 and 0.0 respectively (Table 1). To stable triple helical structure of collagens high percentage of glycine content is needed as integration of large amino acids can reason of steric hindrance [11]**.** Furthermore, proline content was more than 10% in most collagens except collagen alpha-1(XIV) chain precursor and collagen type XII alpha-1 with values 8.8 and 0.0 correspondingly (Table 1). Proline residues of collagen are just as crucial to stabilize the helix and also to act as structural disruptor of the secondary structural elements [12]**.** Therefore, proline concentration is not only helps collagen to act as a structural molecule but also aids in processes like cell-cell adhesion and migration. Other essential physico-chemical parameters were also evaluated through ExPASy's ProtParam (Table 2)**.** The pI values for Alpha1 (XIV) collagen, Alpha-1(V) collagen, Collagen alpha1 CB7, Collagen alpha1 (I) CB6B were found to lie in the alkaline range (pH>7) while for the remaining half, acidic range (pH<7) was observed. Additionally, analysis of instability index classified most collagens as stable (instability index <40) while collagen 4 observed as only unstable (instability index >40) collagen.

Extinction coefficient for Alpha1 (V) collagen and Collagen alpha-1(XIV) chain precursor were monitored higher than remaining members. Besides, higher extinction coefficient indicates higher concentration of lysine, tryptophan and tyrosine residues in the sequence and may be helpful in protein-protein and protein ligand interaction studies in solution. Additionally, Collagen type XII alpha-1 was considered as most thermostable, with highest aliphatic index (126.67) which describing the relative volume of protein occupied by aliphatic side chains assists to study thermo stable properties of an enzyme, followed by Alpha-1(V) collagen (109.40). It is found to span within a range of 31.22 to 126.67. It is a broad range which implies that most of the collagens may be stable. After computing all the members, a wide range of Grand Average of Hydropathy (GRAVY) value was observed from 1.275 to -0.164.

Secondary structural analyses using SOPMA was done for all chicken alpha-1 collagen members and it showed a predominance of random coils, followed by extended strands, α-helices, and β-turns in Collagen alpha-1(XIV) chain precursor while α-helices exceeded the extended strands in, Alpha1 (XIV) collagen, Alpha-1(V) collagen and Collagen type XII alpha-1 (Table 3). High value for random coil allows important significance on packaging of secondary structural elements may which help to derive potential tertiary protein structures, related functions and also promote advancements in protein engineering.

Significant knowledge about the protein's mechanism of action and nature can be get through Motifs of protein. Motif Scan tool used to obtain Signature motifs within protein sequences. Collagen alpha-1(I) chain was described as VWFC domain signature, known to participate in oligomerization, hence forming an imperative part of the complex forming proteins [13, 14].

Phylogenetic tree was built using distance based Neighbor-Joining method. A number of clusters were observed including alpha-1(V) collagen and collagen type XII alpha-1 relating closely to alpha1 (V) collagen, while collagen alpha-1(I)

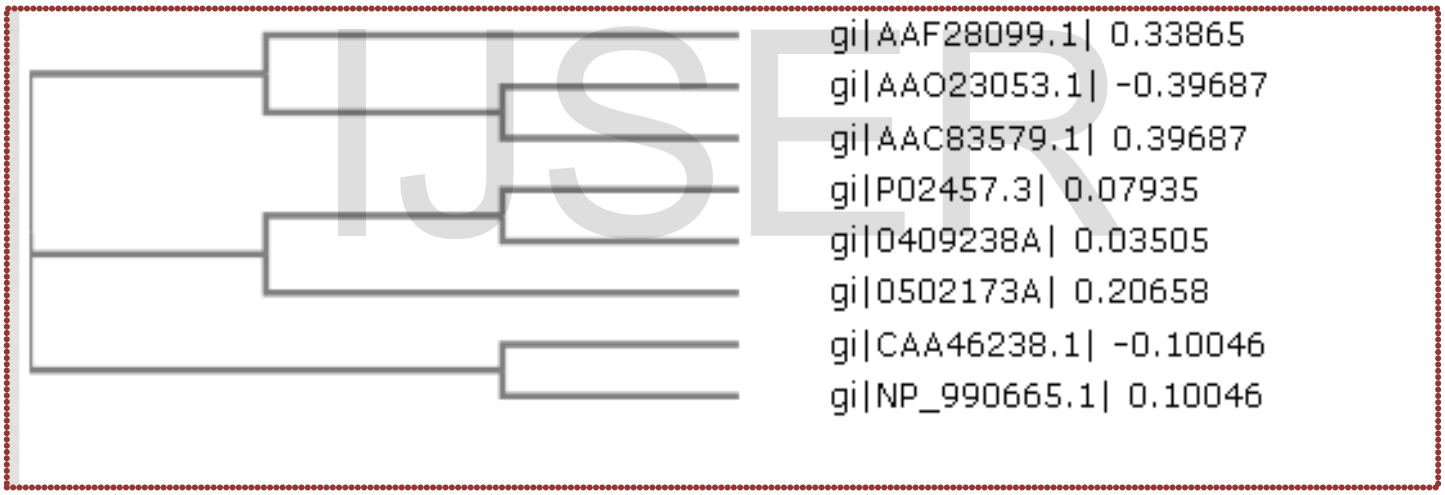

Figure-1: Phylogenetic tree of chicken alpha 1 collagen sequences by using Neighbor-Joining Method

Table 1: Amino acid composition of chicken alpha-1 collagens (in %)

Aln	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ilu	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
5.2	4.0	2.2	5.7	0.7	4.1	6.5	22.9	1.0	3.3	5.0	5.4	1.3	2.2	17.5	3.5	3.8	0.4	2.2	3.2
10.7	4.7	2.1	4.3	1.2	3.1	5.3	26.6	0.6	1.9	3.4	3.9	1.2	1.7	18.7	3.3	3.2	0.4	1.0	2.5
5.8	6.9	2.1	2.1	1.1	7.9	4.2	25.9	1.1	3.2	2.1	1.1	1.6	0.5	19.6	6.3	3.7	0.0	1.1	3.7
5.2	4.0	2.2	5.7	0.7	4.1	6.5	22.9	1.0	3.3	5.0	5.4	1.3	2.2	17.5	3.5	3.8	0.4	2.2	3.2
6.9	4.2	3.3	5.3	1.1	3.9	6.4	10.6	1.4	5.2	7.6	5.1	1.4	3.4	8.8	7.3	7.2	0.8	3.2	6.9
33.3	4.2	0.0	0.0	8.3	0.0	8.3	0.0	0.0	4.2	16.7	0.0	4.2	0.0	0.0	12.5	4.2	0.0	0.0	4.2
12.5	4.8	0.4	3.7	0.0	1.8	3.7	33.2	0.0	1.5	2.2	3.7	0.4	1.1	12.2	2.6	2.6	0.0	0.0	1.5
6.6	4.4	1.1	2.2	0.0	2.2	2.2	33.0	1.1	1.1	4.4	1.1	0.0	2.2	16.5	5.5	1.1	0.0	0.0	2.2

Table 2: Physico-chemical parameters of chicken alpha-1 collagens

Collagen Name	No. of A.A.	No. of A.A.	M.W. (Da)	PI	''_'' Charged residues	''+'' Charged residues	Extinction coefficient	Insta-bility index	Alip-hatic index	Gravy
Alpha1 (V) collagen	1835	1835	184237.0	5.03	223	172	105840	32.57	46.44	-0.864
Collagen alpha-1(I) chain	1453	1453	137754.6	5.47	140	125	56475	28.39	38.52	-0.765
Alpha1 (XIV) collagen	189	189	18273.0	9.30	12	15	3105	39.80	37.20	-0.911
Alpha-1(V) collagen	50	50	5773.7	9.96	5	7	16500	50.38	109.40	-0.164
Collagen alpha-1(XIV) chain precursor	1888	1888	202667.5	5.25	220	176	180140	36.72	77.00	-0.344
Collagen type XII alpha-1	24	24	2365.8	4.53	2	3	125	32.77	126.67	1.275
COLLAGEN ALPHA1 CB7	271	271	24410.0	9.39	20	23	-	-1.12	31.22	-0.611
Collagen alpha1(I) CB6B	91	91	8292.5	8.75	4	5	-	12.17	34.40	-0.503

Table 3: Secondary structural features of bovine alpha-1 collagens (in %)

Collagen Name	α Helix	3_{10} Helix	Pi Helix	β Bridge	Extended Strand	β Turn	Bend Region	Random Coil	Ambigo us States	Other States
Alpha1 (V) collagen	9.70	0.00	0.00	0.00	11.50	4.90	0.00	73.90	0.00	0.00
Collagen alpha-1(I) chain	4.75	0.00	0.00	0.00	7.85	4.75	0.00	82.66	0.00	0.00
Alpha1 (XIV) collagen	16.40	0.00	0.00	0.00	4.76	4.76	0.00	74.07	0.00	0.00
Alpha-1(V) collagen	40.00	0.00	0.00	0.00	8.00	2.00	0.00	50.00	0.00	0.00
Collagen alpha-1(XIV) chain precursor	18.11	0.00	0.00	0.00	21.56	4.08	0.00	56.25	0.00	0.00
Collagen type XII alpha-1	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
COLLAGEN ALPHA1 CB7	0.00	0.00	0.00	0.00	9.23	5.17	0.00	85.61	0.00	0.00
Collagen alpha1(I) CB6B	0.00	0.00	0.00	0.00	1.10	13.19	0.00	85.71	0.00	0.00

IJSER

Chain and collagen alpha1 CB7 closely related to collagen alpha1 (I) CB6B. Moreover, alpha1 (XIV) collagen and collagen alpha-1(XIV) chain precursor has close relation (Figure 1). In close evolutionary relationship proteins may be considered collectively for their association inrelated biological processes. In rapidly growing broiler chickens. The chicken feet collagen is affluent in the amino acids glycine, glutamic acid, proline and hydroxyproline. Electrophoresis pattern confirmed two distinct α-chains (α1 and α2) and β chain which representing that type I collagen is a major component of chicken feet collagen [15]. The composition of amino acid plays an important role in the physical properties of collagen usually glycin is an aboundant amino acid in collagen constitutes around thirty percent of total amino acid content though, the value may differ according to species or body parts [16].

4. CONCLUSION

In this study, we tried to reveal information about chicken alpha-1 collagen by analyzing their structural features e.g. amino acid content, physico-chemical properties, secondary structural features, phylogenetic classification and functional analysis of chicken alpha-1 collagen. A variety of computational tools were used to settle down the process of finding. The findings of this study through may be used by investigators doing on chicken collagens in context of any experimental system. The composition of amino acid demonstrates a considerably high percentage of cysteine residues in collagen 6.

Also, collagen 6 is found to be the most thermo stable collagen. Therefore, we hypothesize that collagen 6 might be a key player in pathological conditions. However, it is known changes in protein structure reasons impairment of protein function and develop many pathological conditions. To validate this proposal more experimental research and testing need to be carried out which will be helpful to analyze certain other groupings and clustering of disease responsive collagens.

5. Acknowledgment

The authors wish to thank Zahida Yesmin Roly in the Dept. of Genetic Engineering and Biotechnology, Faculty of Life and Earth Science, University of Rajshahi, Rajshahi-6205, Bangladesh

6. REFERENCES

[1] Vuorio, E., & De Crombrugghe, B. (1990). "The family of collagen genes. *Annual review of biochemistry*, **59(1):** 837-872".

[2] Hagg, P., Rehn, M., Huhtala, P., Vaisanen, T., Tamminen, M., & Pihlajaniemi, T. (1998). "Type collagen is identified as a plasma membrane protein". *J Biol Chem*, **273:** 15590-15597.

[3] Kadler, K.E., Holmes, D.F., Trotter, J.A., & Chapman, J.A. (1996). "Collagen fibril formation". *Biochem J.*, **316:** 1-11.

[4] Tanzer, M.L. (2006). "Current concepts of extracellular matrix". *J Orthop Sci*. **11**: 326-31.

[5] Bac, I., Osatomy, k., Yoshida, A., Osako, K., Yamaguchi, A., & Hara, K. (2008). "Biochemical properties of Acid-Soluble collagens Extracted from the Skins of Underutilised Fishes". Food Chem., 108:49-54.

[6] Ichie, k., Taguchi, Y., Takahata, Y., Toki, S., Morimatus, F., Yamanaka, T., Kumaga, T., Wataya, Y., Kimura, K., Tanaka, Y., Ikeda, K., Umetsu, A., Shibata, R., Kurisaki, J., & Yamada, R. (1999). "Low allergic gelatin prepared from chicken cartilage by acid processing". 4^{5th} ICoMST, Japan, Proceeding **2:** 714-715.

[7] Saha, A., Alam, M.J., Ashraf, K.U., & Mannan, A. (2013). "Computational analysis of bovine alpha-1 collagen sequences". *Bioinformation,* **9(1):** 42-8.

[8] Islam, M.M., Haque, M.E., Dutta, A.K., Islam, M.A., Khalekuzzaman, M., & Sikdar, B. (2013). "A Computational Study on Metalloproteinase in the Snake Venomous of Different Rattle Snake Species in Comparison with the Asian Snake Species". *International Journal of Pharma & Bio Sciences*, 4(4): 741 – 748.

[9] Roly, Z.Y., Islam, M.M., & Reza, M.A. (2014). "A comparative in silico characterization of functional and physicochemical properties of 3FTx (three finger toxin) proteins from four venomous snakes". *Bioinformation,* **10(5):** 281-7. [PMID: 24966535]

[10] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., & Bairoch, A. (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis". *Nucleic Acids Res.* **31(13):** 3784-8.

[11] Van der Rest M., & Garrone, R. (1991). "Collagen family of proteins". *FASEB J.*, **5(13):** 2814-23.

[12] Ramachandran, G.N., Bansal, M., & Bhatnagar, R.S. (1973). "A hypothesis on the role of hydroxyproline in stabilizing collagen structure". *Biochemica et Biophysica Acta*., **332**: 166-71.

[13] Hunt, L.T., & Barker, W.C. (1987). "Von Willebrand factor shares a distinctive cysteine-rich domain with thrombospondin and procollagen". *Biochem Biophys Res Commun,* **144**: 876-82.

[14] Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., Van Mourik, J.A., & Pannekoek, H. (1991). "Assembly and routing of von Willebrand factor variants: the requirements for disulfide-linked dimerization reside within the carboxy-terminal 151 amino acids. Assembly and routing of von Willebrand factor variants: the requirements for disulfide-linked dimerization reside within the carboxy-terminal 151 amino acids". *J Cell Biol.,* **113:** 195-205.

[15] Hashim, P.M.S., Mohd Ridzwan, & Bakar, J. (2014). "Isolation and Characterization of Collagen from chicken Feet". *International journal of Agricultural, Biosystems science and Engineering*, **8(3):** 238-242.

[16] Gimenez, B., Turnay, J., Lizarbe, M.A., Monetro, P., & Gornez Guillen, M.A. (2001). "Use of Lactic Acid for Extraction of Fish and its Characteristics". *Asian Aust. J. Animal Sci.,* **14:** 1638-1644.

IJSER