Research Article

Characterization and protein modelling of buffalo S100A8 reveals conserved molecular structure in bovines

Manika Sahgal, S. Goyal, P.K. Dubey, R.S. Kataria and S.K. Niranjan*

ICAR-National Bureau of Animal Genetic Resources, Karnal-132001, Haryana, India

ABSTRACT

S100A8, a calcium binding protein, is involved in various intracellular and extracellular functions. In this paper, we characterized the complete cDNA sequence of the S100A8 molecule and modelled its protein structure in buffalo (*Bubalus bubalis*). The structure was further compared with that of cattle to determine molecular homology. The amino acid sequence of buffalo S100A8 was 89 residues long and showed 12 amino acid variations in comparison to cattle. Sequence divergence of buffalo S100A8 with non-ruminants was higher than that of ruminant species. Phylogenetic analysis revealed that buffalo and other ruminants were found to group together separated from all other non-ruminants, including human. Physiochemical parameters of buffalo S100A8 showed comparatively higher stability as well as structurally conserved protein structure. Comparative modelling, using cattle as a reference, confirmed the high structural conservation of buffalo S100A8. Further, amino acid substitutions at seven important locations of buffalo S100A8 (7, 11, 22, 23, 36, 68 and 81) did not appear to affect the protein model with that of cattle. This study highlights the molecular structure of S100A8 in buffalo, demonstrating high sequence homology and conserved protein architecture among bovines.

Key words: *Bubalus bubalis*, Protein modelling, S100A8, Sequence analysis

***Corresponding author:** saketniranjan@gmail.com

INTRODUCTION

S100 are low molecular weight calcium binding proteins, having both intracellular and extracellular functions including regulating intracellular transduction/ phosphorylation pathways, membrane, cytoskeletal interactions energy metabolism, cell migration and pro-inflammatory functions (Newton and Hogg, 1998; Zimmer *et al.,* 2003; Mackay *et al.,* 2003; Santamaria-Kisiel *et al.,* 2006). Although the S100 proteins are structurally related, differential expression in general as well as in various disease conditions has also been observed among them (Sheng *et al.*, 1994; Abraha *et al.,* 1999; Marenholz *et al.,* 2004; Tirkos *et al.,* 2006).

The S100A8, also known as Migration inhibitory factor related protein-8 (MRP8) or calgranulin A, is a myeloid associated S100 protein, constitutively expressed in neutrophils. Along with S100A9, it forms S100A8/A9 complex (caprotectin), which has been found to have anti-proliferative and antimicrobial activities. During inflammatory process, the S100A8 positive neutrophils appear very early at the site of inflammation and function as an inflammatory marker protein (Johne *et al.,* 1997, Passey *et al.,* 1999). In dairy cattle, differential expression of this protein has been reported during inflammation of mammary gland (Ogorevc *et al.,* 2008).

The S100A8 protein is encoded by *S100A8* gene, a member of multi gene family. The *S100A8* gene has been extensively studied in many mammalian species, however, scanty information is available in livestock species. In the present study, cDNA sequence of *S100A8* gene has been characterized in water buffalo (*Bubalus bubalis*). Further, protein model of the S100A8 was also evaluated and compared with that of cattle to assess molecular characteristics and structural conservation.

MATERIALS AND METHODS

cDNA amplification of buffalo S100A8

Buffalo mammary gland tissue sample was collected during routine slaughtering from one healthy adult female after cessation of milk production. Portions of each tissue were removed immediately after slaughtering and preserved in RNALater (Sigma-Aldrich) and transported to the laboratory. Finally, tissues were stored at -80°C for further use after removing RNALater completely. Total RNA was isolated from approximately 250mg of homogenized tissue, by TRIzol method (Invitrogen) according to manufacturer's protocol. The extracted RNA was further purified and treated with *DNaseI* using MinElute RNA cleanup kit (Qiagen) as per manufacturer's instructions to remove genomic DNA contamination. Purified RNA was quantified by UV spectrophotometer (NanoDrop, ND-1000). Purity of RNA was confirmed by OD 260/OD 280 ratio, which was 1.9–2.0 for all the samples. Reverse transcription was performed using Oligo dT primers using 2 µg of the total RNA from each sample using RevertAid™ first strand cDNA kit (MBI Fermentas) following manufacturer's instructions. RNA

sample was denatured at 65˚C for 5 min, snap cooled on ice and reverse transcribed at 42˚C for 60 min. The reverse transcription step was stopped by incubating at 70˚C for 5 min. Two microliter of the cDNA was used as template for PCR amplification.

One pair of oligonucleotide primers (Forward - 5'- TATTTTGGGGAGACCTGGTG-3' and Reverse 5'-ACTCTGCAGGCACGAGTCC-3') was designed from cattle sequence (GenBank Acc. No. NM_001040470, Table1) using Primer3 programme (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) to amplify the complete cDNA of buffalo *S100A8* gene. PCR was carried out in 25 ml volume with final concentration of 1.5 mM $MgCl_2$, 200 mM dNTPs, 10 pmol primers each (forward and reverse), 1X PCR buffer and 1U *Taq* polymerase (MBI Fermentas). Amplification conditions were: Initial denaturation at 94°C for 3 min, thirty cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec and extension at 72°C for 1 min and final extension at 72°C for 10 min. After ascertaining the amplification of the desired product by gel electrophoresis in 1.5% agarose gel, the PCR products were sequenced from both directions by automated sequencing method. DNA sequencing of both the PCR products representing complete coding region of buffalo *S100A8* gene was carried out by using single-pass on ABI PRISM® 3100 Genetic Analyzer. Annotated sequence of complete *S100A8* gene was further submitted to NCBI GenBank with Acc. No. KF511591. Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast) was used to estimate the homology between buffalo S100A8 and sequences of other mammalian species. Sequence alignment analysis and domain structure predictions were carried out by UNIPROTKB (http://www.uniprot.org/blast/uniprot/).

Phylogenetic analysis of mammalian S100A8

The evolutionary history of mammalian S100A8 based on complete cDNA and amino acid sequences was inferred by Neighbor-Joining method using MEGA5 programme. To construct phylogenetic tree, the sequences were aligned using Multiple Sequence Alignment (MUSCLE), and evolutionary distances were computed using the p-distance method. Bootstrap values were also obtained using 500 replicates.

Protein structure prediction

For functional characterization of S100A8 in *Bos taurus* and *Bubalus bubalis*, the 3D protein structures were modelled and were then subjected to further analysis. To thoroughly comprehend the differences among S100A8 protein sequences of *Bubalus bubalis* and *Bos taurus* at structural level, the structure was modeled via homology modeling using 1mr8.1.A i.e. migration inhibitory factor-related protein 8 from human as a template.

Physiochemical properties were computed using ProtParam tool. The structure was determined by aligning the protein sequence of S100A8 with the template of the available structure and loop refinement. The model was evaluated using diverse statistical constraints and parameters such as Qualitative Model Energy ANalysis (QMEAN4) and Global Model Quality Estimation (GMQE). The two structures were then superimposed for comparison, facilitating the analysis of large biomolecular 3D systems through an interactive graphical interface and integrated scripting tools (Humphrey et al, 1996).

RESULTS AND DISCUSSION

A 331 bp long fragment of buffalo *S100A8* cDNA was sequenced, annotated and submitted to NCBI-GenBank (accession number KF511591). A 270 nucleotides long Open reading frame (ORF), encoding 89 amino acid residues of S100A8 protein molecule was positioned between 26 to 295 nucleotides. The length of the ORF was found to be similar to that of other species.

Evolutionary divergence and phylogenetic analysis

Nucleotide identity based on ORF regions of buffalo S100A8 was 96, 94, 86, 86, 85, 84 and 81% with that of taurine cattle (Acc. No. NM_001113725), sheep (Acc. No. XM_004002523), pig (Acc. No. AK399753), horse (Acc. No. XM_001493589), cat (Acc. No. XM_003999782), dog (Acc. No. NM_001146144) and human (Acc. No. NM_002964), respectively. As an estimate of evolutionary divergence between cattle and buffalo S100A8, the number of base substitutions per site was 0.044, which appears to be higher compared to the divergence observed for most of the conserved genes between these species. The evolutionary divergence of buffalo *S100A8* at nucleotide level was found to be significantly higher (upto three times) than that of non-ruminant species when compared to other ruminant species. Low divergence among ruminants largely separate this group from other species, indicating specificity of S100A8 molecule in these species. Buffalo S100A8 showed a total of 12 nucleotide variations compared to taurine cattle, out of which seven nucleotides were found to be non-synonymous types. The amino acid changes were observed at 7, 11. 22. 23. 36. 68 and 81 residue positions (Fig. 1).

Phylogenetic analysis based on *S100A8* cDNA sequence revealed clustering of ruminant species (Fig. 2). The diversification of *S100A8* genes in ruminants appears to be a relatively recent event. However, the lineage of *S100A8* gene in ruminants diverged much earlier from common ancestors of ruminants and non-ruminants. It

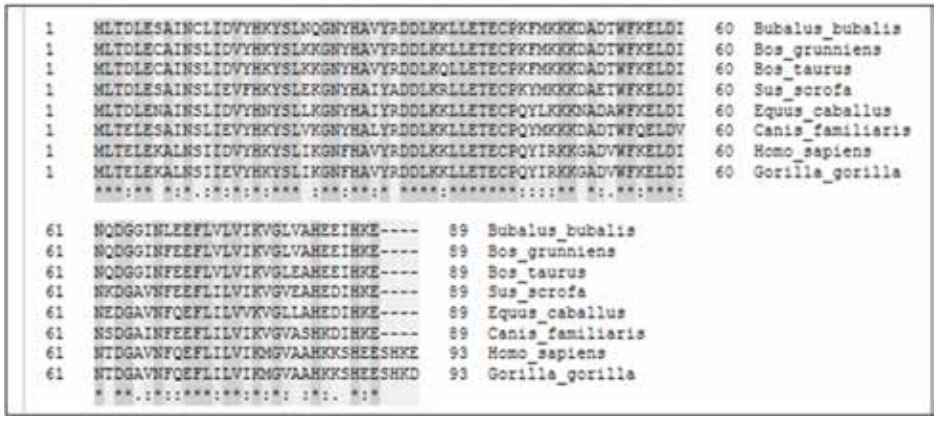

Fig 1: *Multiple alignment of conceptualized amino acid sequence of S100A8 gene (based on cDNA sequence of different species).*

has been reported that S100 protein coding genes are found in only primates, suggesting that they belong to a young gene family. The recent diversification of S100A8 in ruminants, however, indicates that this gene family is still under evolutionary selection.

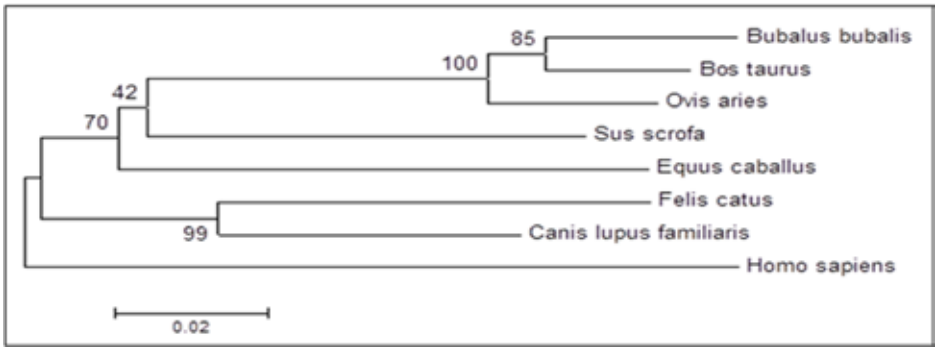

Fig 2*: cDNA sequence based phylogenetic analysis of S100A8 gene among various species.*

Protein sequence analysis and structure prediction

Physiochemical properties computed using ProtParam tool, indicated that the molecular weight of the buffalo S100A8 protein was 10381.9 dalton, with atheoretical pI of 5.0. The protein contains Total number of 11positively and 17negatively charged residues. Its instability index is 26.74, classifying it as a stable protein, with an estimated half-life of approximately 30 hrs for mammalian reticulocytes, in vitro. The Grand average of hydropathicity (GRAVY) was estimated to be (-) 0.349. The extinction coefficient was calculated to be 11585 $M^{-1}cm^{-1}$. The differences among *Bos taurus* and *Bubalus bubalis* S100A8 protein sequences at structural level were determined by homology modelling using migration inhibitory factor-related protein 8 (1mr8.1.A) template from human in PDB database. This sequence shared 69.66% sequence identity with S100A8 (*Bubalus bubalis*) and 71.91% sequence identity with S100A8 (*Bos taurus*). For evaluation of model using diverse statistical constraints and parameters, the QMEAN4 and GMQE values for buffalo were (-) 1.21 and 0.87 respectively, compared to -1.31 and 0.89 for cattle.

The 3D protein structures were modelled for functional characterization of S100A8 in *Bos taurus* and *Bubalus bubalis* (Fig. 3). Both models showed exact homology, showing complete conservation of secondary protein structures of alpha helixes and beta sheets. Superimposing of the protein models also indicated conserved structure of both species (Fig. 3 c). Amino acid substitutions at seven important locations of buffalo S100A8 protein model (7, 11, 22, 23, 36, 68 and 81) did not appear to affect the secondary structure (Fig. 4). Amino acids at six positions were replaced with the amino acids of similar properties and affinities to water (hydrophilic/ hydrophobic). Importantly, substitution of hydrophilic glutamine with hydrophobic valine of at residue position 81 did not influence the alpha helix structure, likely due to the position of the residue at end region.

In conclusion, cDNA sequence of S100A8 molecule was characterized in riverine buffalo, revealing seven amino acid differences compared to cattle. Phylogenetic analysis based on cDNA sequences revealed that buffalo and other ruminants grouped together, distinct from non-ruminants. Further, amino acid sequence

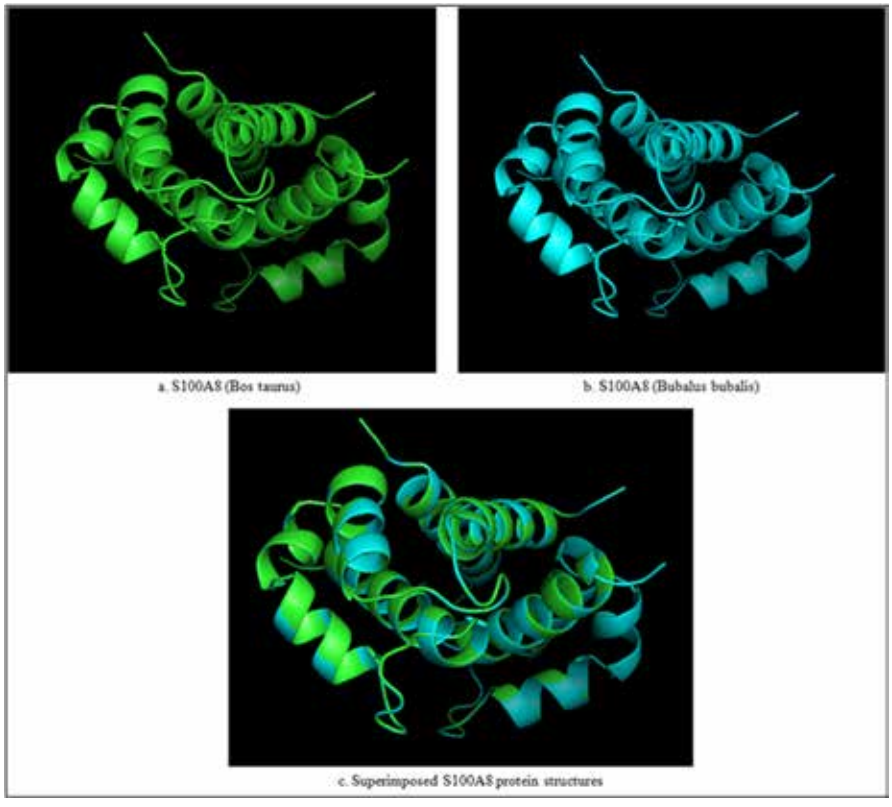

Fig 3: *Protein structure of S100A8 (a) cattle, (b) buffalo and (c) superimposed protein structures of cattle and buffalo.*

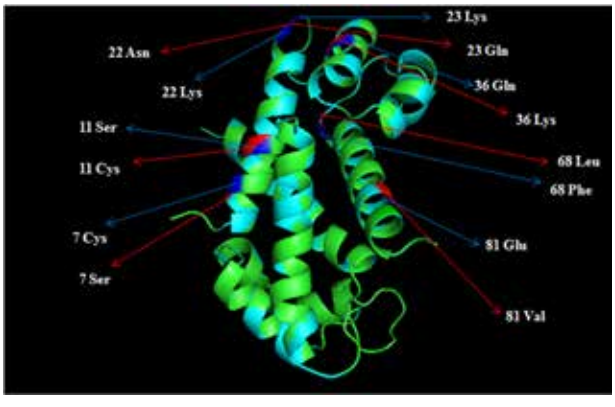

Fig 4: *Important residue sites with amino acid substitutions in superimposed S100A8 protein structure of cattle and buffalo*

homology as well as protein modelling revealed high homology and conserved protein structure of S100A8 molecule among bovines.

ACKNOWLEDGEMENT

The authors are thankful to the Director, ICAR-NBAGR, Karnal and Indian Council of Agricultural Research for providing facilities to carry out the present study. Technical assistance of Mr. Naresh Kumar is gratefully acknowledged.

REFERENCES

Abraha HD, Noble PL, Nicolaides KH, and Sherwood RA. 1999. Maternal serum S100 protein in normal and Down syndrome pregnancies. *Prenat. Diagn*, 19:334-336.

Humphrey W, Dalke A, and Schulten K. 1996. VMD: visual molecular dynamics. *J. Mol. Graph,* 14(1):33-8.

Johne B, Fagerhol MK, Lyberg T, Prydz H, Brandtzaeg P, Naess-Andresen CF, and Dale I. 1997. Functional and clinical aspects of the myelomonocyte protein calprotectin. *J. Clin. Pathol. Mol. Pathol,* 50:113-123.

Mackay A, Jones C, Dexter T, Silva RL, Bulmer K, Jones A, Simpson P, Harris RA, Jat PS, Neville AM, Reiss LF, Lakhani SR, and O'Hare MJ. 2003. cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene,* 22:2680-2688.

Marenholz I, Heizmann CW, and Fritz G. 2004. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem. Biophys. Res. Commun,* 322:1111-1122.

Newton RA, and Hogg N. 1998. The human S100 protein MRP-14 is a novel activator of the beta 2 integrin Mac-1 on neutrophils. *J. Immunol,* 160:1427-1435.

Ogorevc J, Kunej T, and Dovc P. 2008. An integrated map of cattle candidate genes for mastitis: a step forward to new genetic markers. *Acta Agriculturae Slovenica, Supplement,* 2:85-91.

Passey RJ, Xu K, Hume DA, and Geczy CL. 1999. S100A8: emerging functions and regulation. *J. Leuk. Biol,* 66:549-556.

Santamaria-Kisiel L, Rintala-Dempsey AC, and Shaw GS. 2006. Calcium dependent and independent interactions of the S100 protein family. *Biochem. J,* 396:201-214.

Sheng JG, Mrak RE, and Griffin WS. 1994. S100 beta protein expression in Alzheimer disease: potential role in the pathogenesis of neuritic plaques. *J. Neurosci. Res,* 39:398-404.

Tirkos S, Newbigging S, Nguyen V, Keet M, Ackerley C, Kent G, and Rozmahel RF. 2006. Expression of S100A8 correlates with inflammatory lung disease in congenic mice deficient of the cystic fibrosis transmembrane conductance regulator. *Respir. Res,* 7:51.

Zimmer DB, Wright Sadosky P, and Weber DJ. 2003. Molecular mechanisms of S-100 target protein interactions. *Microsc. Res. Tech*, 60:552-559.