Characterization and in silico Analysis of Pregnancy Associated Glycoprotein-1 Gene of Buffalo (Bubalus bubalis)

1A. Jerome, 2S.K. Singh, 3S.K. Agarwal, 4Mohini Saini and 5Ashwin Raut
1Division of Animal Reproduction, Indian Veterinary Research Institute, Izatnagar, Bareilly 243 122 (U.P.), India
2Centre for Wildlife, Indian Veterinary Research Institute, Izatnagar, Bareilly 243 122 (U.P.), India
3Division of Animal Biotechnology, Indian Veterinary Research Institute, Izatnagar, Bareilly 243 122 (U.P.), India

Abstract: Pregnancy Associated Glycoproteins are trophoblastic proteins belonging to the Aspartic proteinase family secreted by different mammalian species. They play a pivotal role in placentogenesis, foeto-maternal unit remodeling and implantation. The identification of the genes encoding those proteins will be helpful to unravel the intricate embryogenomic functions during pregnancy establishment. Considering importance of these proteins, the present study was undertaken to characterize the pregnancy associated glycoprotein-1 gene of buffalo. An 1181 base pairs buffalo Pregnancy Associated Glycoprotein PAG-1 gene was PCR amplified from the RNA obtained from the fetal cotyledons. BLAST analysis of the buffalo PAG-1 sequence retrieved a total of 94 sequences which includes 35 cattle, 5 goat and 4 sheep PAG sequences, exhibiting more than 80% similarity. In silico analysis of buffalo PAG-1 gene revealed an uninterrupted open reading frame of 1140 base pairs encoding 380 amino acids possessing 15 amino acid signal peptide and mature peptide of 365 amino acids which is proteolytically inactive due to key mutations. The potential sites for glycosylation were also deduced. The phylogenetic study of the buffalo PAG-1 gene revealed buffalo PAG-1 is more related to cattle, goat and sheep PAG-1 sequences. Characterization and phylogenetic analysis of buffalo PAG-1 gene revealed many conserved regions and its proteolytic inactivity.

Keywords: Buffalo, Bubalus bubalis, pregnancy, Pregnancy Associated Glycoprotein (PAG), aspartic proteinase, phylogenetic

INTRODUCTION

Pregnancy is established and maintained by the two-way communication between the conceptus and the mother. These intricate dialogues which are initiated post fertilization are crucial as these signals are considered potential markers for effective placental remodeling, pregnancy recognition and successful implantation. These interactions between the conceptus and maternal system emphasize the importance of both the components in maternal recognition of pregnancy and embryonic development (Roberts et al., 2008).

These important signals to the maternal system to sustain pregnancy are mediated by numerous molecules which include steroid hormones, peptide hormones, cytokines and growth factors (Roberts et al., 2008; Salamonsen, 1999).

Conceptus derived substances are considered to be precise and reliable markers of pregnancy and fetal well being. Pregnancy associated glycoproteins are one such large family of protein molecules produced by conceptus for the recognition by the mother. Pregnancy associated glycoproteins are acidic glycoprotein belonging to the Aspartic Proteinase superfamily sharing >50% amino acid sequence identity with Pepsin, Cathepsin D and E (Sousa et al., 2006; Xie et al., 1991).

Pregnancy associated glycoproteins form very large family of glycoproteins; nearly 22 different PAGs in ruminants have been identified at different stages of gestation (Xie et al., 1994). Pregnancy Associated Glycoprotein-1 (PAG-1) also known as Pregnancy Specific Protein B (PSPB), FSP-60 and SBU3 is secreted by the binucleate cells of the conceptus trophoderm (Xie et al., 1997).

Corresponding Author: A. Jerome, Buffalo Physiology and Reproduction Division, Central Institute or Research on Buffaloes, Hisar, 125001, Haryana, India
PAG-1 is detectable in maternal blood soon after implantation as binucleate cells migrate from the trophoectoderm and fuse with uterine epithelial cells and hence it is considered as a potential signal from the conceptus (Garbayo et al., 2008). The products of binucleate cells in maternal circulation have also been reported to be associated with placental mass, fetal number, twins and neonatal birth weight in cattle (Butler et al., 1982; Zoli et al., 1991). The PAGs are multigene family expressed in placenta of eutherian mammals and their expression varies spatially as well as temporally during gestation (Klisch et al., 2005). Multiple PAG genes have been cloned and identified in many domestic animals such as cattle, sheep (Xie et al., 1994), goat (Sasser et al., 1986), pig (Patel et al., 1995) and wild ruminants’ species (Xie et al., 1994). Based on the evolutionary study and phylogenetic linkage bovine pregnancy associated glycoproteins family has been segregated as ancient (bovine PAG-2, bovine PAG-8) and modern (bovine PAG-1) (Vasques et al., 1995, Garbayo et al., 2000, Szafrianska et al., 1995). But there is no report on the characterization and phylogenetic analysis of pregnancy associated glycoprotein-1 gene of buffalo. Moreover, identification of gene encoding buffalo Pregnancy Associated Glycoprotein-1 may provide an avenue for producing recombinant protein which will be helpful to develop diagnostics for early pregnancy diagnosis and marker for embryonic development (Rawan and Cross, 2008). Considering the importance of the gene in embryogenesis, the present study was designed to characterize pregnancy associated glycoprotein-1 (PAG-1) gene and analyze its phylogenetic lineage.

MATERIALS AND METHODS

Sample collection and RNA isolation: Buffalo placenta were collected from local abattoir. The stage of pregnancy was estimated by measurement of crown-rump length. Total RNA was isolated from fetal cotyledons using TRI reagent (Ambion, USA) following manufacturer’s instructions. The integrity of the extracted RNA was checked by agarose gel (1%) electrophoresis and visualization of the gel under UV light after staining with ethidium bromide. The purity of the obtained RNA was checked by means of spectrophotometric readings at OD_{260}/OD_{280}.

cDNA synthesis and confirmation of cDNA PCR buffalo PAG-1 gene amplification: RNA from the fetal cotyledons was reverse-transcribed into cDNA with reverse transcriptase (Qiagen, Germany), oligo (dT) primers and 500 μM dNTPs at 37°C for 1 h. On the basis of available PAG-1 sequences from cattle (GenBank Acc. No. M73662; NM 174411), goat (GenBank Acc.No. AF191326), sheep (GenBank Acc.No. M73961), pig (GenBank Acc. No. L34360), white-tailed deer (GenBank Acc. No. AY508665) and zebra (GenBank Acc.No. AF036952) buffalo gene specific primers were designed using Primer select programme of DNA star software. The primers were forward (PAG-1/start F, 5'-GG AT CC AGG AAA TAA A CATGAA GTG-3' and PAG-1/stopR, 5'-TTACTG AAC CACTCYMAGCATT-3').

PCR amplification was carried out in a total volume of 25 μL of reaction mixture containing approximately 100 ng of cDNA, 10X PCR buffer (100 Mn Tris-Hcl, pH 8.8 at 25°C, 5 pM of forward and reverse primer of each, 2.0 mM MgCl2, 200 μM dNTPs, 1.0 U Taq DNA Polymerase. The PCR protocol involved an initial denaturation at 94°C for 2 min; 30 cycles of denaturation (94°C for 15 sec), annealing (optimum temperature of 51.3°C for 15 sec) and extension (74°C for 45 sec) and one cycle of final extension (74°C for 10 min). The PCR product was checked by 1% agarose gel electrophoresis.

cDNA cloning and sequencing: PCR amplicons were cloned in cloning vector (PTZ57R/T in sTaconic, MBI, Fermentas) following the manufacturer’s protocol. The 12 μL of ligated product was directly added to 200 μL competent cells and cells were then immediately transferred on chilled ice for 5 min and SOC was added. The bacterial culture was pelleted and plated on LB agar plate containing Ampicillin (100 μg mL⁻¹) added to agarplate @1: 1000, IPTG (200 mg mL⁻¹) and X-Gal (20 mg mL⁻¹) for blue-white screening. Isolation of plasmid from was done using kit’s protocol (Biochem Life sciences). Recombinant plasmids were characterized by PCR using gene-specific primers and restriction enzyme digestion.

Restriction enzyme analysis of the plasmid was carried out with Xba I (MBI, Fermentas USA) and Sma I (MBI, Fermentas USA) enzymes.

Sequencing and sequence analysis of buffalo PAG-1 gene: The plasmid containing buffalo PAG-1 cDNA was sequenced using M13 Forward and Reverse primer pair by primer walking in an automated DNA sequencer (Sequence Analyzer Version 2.0, ABI Prism, Chromous Biotech, Bangalore). The sequence obtained was subjected to BLAST (www.ncbi.nlm.nih.gov/BLAST) to retrieve PAG-1 sequences of other species. The nucleotide and deduced amino acid sequences of buffalo PAG-1 gene were aligned and compared with other species sequences available in GenBank using Clustal
option in MegAlign (Lasergene Software, DNASTAR). The buffalo PAG-1 protein structure was predicted by online SWISS MODEL software. The domain structure, glycosylation sites and hairpin loop structure were determined by online software like, PROSITE (http://www.expasy.ch/pro), SMART (http://smart.embl-heidelberg.de/).

The predicted buffalo PAG-1 protein sequence was statistically analysed using SAPS software. Phylogenetic tree based on the evolutionary distances was constructed using MegAlign (Lasergene Software, DNASTAR), based on the nucleic acid and amino acid alignment. Using MEGA 4.1 software the number of synonymous substitution per synonymous site (dS) and number of non-synonymous substitution per nonsynonymous sites (dN) were estimated and neutral (dS = dN), positive (dN>dS) or purifying (dN<dS) selections were tested with a codon-based Z-test using Nei Gojobori method.

**RESULTS AND DISCUSSION**

**Buffalo PAG-1 gene:** The concentration of RNA was checked by analyzing OD_{260}/OD_{280} ratio which was found in the range of 1.8-1.9 indicated the purity of the RNA and the yield was obtained in range of 3.5-3.8 μg mL^{-1}. From the obtained total RNA, cDNA was synthesized and PCR amplification was carried out in 1% agarose gel. Agarose gel electrophoresis revealed 1811 bp PCR product of buffalo PAG-1 gene (Fig. 1). Following this restriction enzyme analysis of the plasmid with Xba I and Sma I enzymes released an insert of 1181 bp.

**Sequence analysis:** BLAST analysis of the amplicon sequence retrieved 35, 5 and 4 sequences of bovine, caprine and ovine PAG, respectively that exhibited more than 80% similarity with buffalo PAG-1 sequence and 5 sequences of white tailed deer which showed less than 80% homology. Accordingly, the sequence was submitted to Genbank as buffalo Pregnancy Associated Glycoprotein-1 gene under accession number EU815059. The deduced buffalo PAG-1 gene consists of an open reading frame of 1140 nucleotide corresponding to an inferred polypeptide length of 380 amino. Moreover, on translation buffalo PAG-1 gene sequence encodes a signal sequence constituting the first 15 amino acids and a mature peptide of 365 residues.

With the members of aspartic proteinase family buffalo PAG-1 sequence showed a similarity of 62.4, 34.6, 33.8, 32.6, 29.5 and 28.7% with pepsinogen, cathepsin E, chymosin, pepsin, cathepsin D and rennin respectively at the amino acid level. With cattle PAG gene sequences buffalo PAG-1 sequence showed a similarity >80% with PAG-1, 3, 16 and <80% with other cattle PAGs.

Fig. 1: Agarose gel electrophoresis depicting PCR amplified Buffalo PAG-1 gene; Lane 1 and 2: 1181 bp buffalo PAG-1 gene; Lane M: 100 bp DNA ladder as molecular size marker

Buffalo PAG-1 sequence showed highest similarity of 89.3% with cattle PAG-1 and lowest of 54.0% with cattle PAG-8.

**Phylogenetic analysis:** The phylogenetic analysis of the obtained sequence of buffalo PAG-1 gene (GenBank Acc No. EU815059) with members with aspartic proteinase family members revealed buffalo PAG-1 sequence is more related to pepsinogen and less related with rennin (Fig. 2). The degree of similarity was determined comparing the obtained buffalo nucleotide and derived amino acid sequence with available PAG-1 sequences of cattle, goat, sheep, horse, zebra, white tailed deer, pig and cat. Buffalo PAG-1 showed the percent similarity ranging from 47-89% other species PAG-1 sequences. Buffalo PAG-1 sequence showed highest similarity of 89.2% with cattle when compared to other species.

Phylograms of buffalo PAG-1 sequence with other species showed buffalo, cattle, sheep, goat, horse, white tailed deer, zebra, feline comprise one clade and pig comprise another. Perceive PAG-1 stands alone and represents an entirely different clade. The PAG-1 sequence of cattle and buffalo belonged to same group and showed more closeness to goat. The PAG-1 of horse, deer and zebra seems to be an evolutionary connecting link between ruminants and other non ruminants. Horse, deer and zebra show more evolutionary closeness to each other but far distant from ruminants (Fig. 3). Analysis of
Fig. 2: Phylogram depicting the evolutionary relationship of buffalo PAG-1 with members of aspartic proteinase family based amino acid sequence

Fig. 3: Phylogram depicting the evolutionary relationship of buffalo PAG-1 with PAG-1 gene sequences of other species based amino acid sequence

phylogenetic tree constructed with buffalo PAG-1 and cattle PAG sequences revealed buffalo PAG-1 gene is more closely related to cattle PAG-1, 3, 21 and 19 both at nucleotides and amino acids level. Cattle PAG-2 group occupies an intermediate position between the ancient aspartic proteinase family members and PAG-1 group. Buffalo PAG-1 as well as other species PAG-1 sequences comprise one clade and show more divergence from the ancient aspartic proteinase family members (Fig. 4). Codon-based Z-test using the Nei Gojobori method revealed that at 5% level of significance, dN is substantially greater than dS. Thus, buffalo PAG-1 might have evolved by recently by positive selection (dN>dS) among these species. Buffalo and cattle PAG-1 gene showed identical lineage. Goat, sheep, horse and deer are similar but have different lineage. Pig sequence show dissimilarities suggesting different ancestry.

**Predicted protein analysis:** The predicted buffalo PAG-1 protein sequence was statistical analyzed using SAPS software revealed the total number of negatively and positively charged residues charged residues were 30 and 40, respectively. The atomic composition of the protein Consists of carbon (C) 1970; Hydrogen (H) 3011; Nitrogen (N) 521; Oxygen (O) 539; Sulfur (S) 15 with 6056 atoms. The extinction coefficient is 1.828/M/cm, at 280 nm measured in water. The aliphatic index and theoretical pl is 93.32 and 9.10, respectively with hydrophoticity index of 0.071. The estimated half-life of buffalo PAG-1 protein is 30 h with an instability index of 32.12 predicting the protein as stable.

Sequence of deduced PAG-1 gene of buffalo revealed an insert of 1181 bp. Comparing the buffalo nucleotides and derived amino acids with available PAG-1 sequences of other species confirm buffalo PAG-1 contains an open reading frame of 1140 nucleotides which on translation corresponds to a polypeptide length of 380 amino acids similar to cattle. On translation, buffalo PAG-1 cDNA yields a polypeptide of 380 amino acids with a signal peptide encoded by the first 15 amino acids followed by a mature peptide of rest 365 residues. The signal sequence of the derived buffalo PAG-1 amino acid sequence (MKWLVLLLGLVAFSEC) starts and terminates with methionine and cysteine, respectively. This signal sequence is well conserved in buffalo as in bovine and other species (Sousa et al., 2006) (Fig. 5). The predicted buffalo PAG-1 protein structure along with its amino acid residues was also deduced (Fig. 6).

The derived mature peptide sequence consists of well conserved motif (ISF1RGS) between the pro-peptide and the mature molecule. The arrow indicates the site of cleavage between the propeptide and the mature protein.

The N-terminal ends of ruminant PAG native proteins predominantly contains the sequence of 3 amino acids, 1-R-D/G-S which is also deduced in buffalo PAG-1 derived amino acid sequence (Telugu et al., 2009; Patel et al., 2004). On comparison of the amino acid
Fig. 5: MegAlign report of buffalo PAG-1 amino acid sequence (380 amino acids) with PAG-1 amino acid sequences of other species
sequences of buffalo PAG-1 with other species PAG sequences, it is evident that both amino terminus 91-98 (VVFDTGSS) and carboxyl terminus 278-284 (LVDTGTS) is conserved across the domestic species (90-100%) suggesting importance of this region for the diagnostic function of PAG-1 and suggesting that the buffalo PAG-1 has also evolved from same family (Sousa et al., 2006; Hughes et al., 2000). Moreover, residues flanking the aspartic acid residues which are considered to be essential for catalytic activity of pepsin are well conserved in buffalo as in other domestic species (Green et al., 1998).

Although, the PAGs clearly belong to the aspartic proteinas, they are not active proteolytically owing to the key mutations close to the active site due to amino and carboxyl terminal flanking the aspartic acid residues that would likely interfere with the catalytic mechanism. On analysis it was evident that buffalo PAG-1 is not proteolytically active since it possesses key mutation of alanine substitution in place of glycine at the active site as same as boPAG-1 which will displaces a water molecule that normally resides between the two catalytic aspartic acids and is directly involved in the catalytic mechanism (Sousa et al., 2006; Green et al., 1998). Although PAG-1 group molecules are not proteolytically active, they have retained the characteristic bilobed structure of aspartic proteinas (Green et al., 1998).

The catalytic mechanism which leads to peptide bond cleavage in the middle of the bound peptide is initiated through nucleophile attack by a hydroxyl ion supplied by a water molecule strategically positioned between the two aspartic acids (Hughes et al., 2000). Comparison of buffalo PAG sequence and other species sequence reveals segments of primary structure that are hypervariable and others that are relatively constant. The region between residues 70 and 100 (13-42 pepsin numbering) showed very few substitutions depicting the presence of other conserved sequence (Xie et al., 1991). From the alignment study it was evident that buffalo PAG-1 which showed more similarity to cattle PAG-1 will also possess a three-dimensional fold and may have a general affinity for binding peptides with basic residues (Hughes et al., 2000).

The deduced buffalo PAG-1 amino acid sequence contains microsequences YS (position 46, 47), LSQSF (position 48-58), RGSNLTH (position 59-66), PLRN (position 67-70) and IKDLVYMGNITGTP (position 71-81) which are found to be conserved across ruminant PAG proteins.

In general, the conserved regions are ones that are internal and structurally important for retaining the overall three-dimensional fold of the molecule. By contrast, the hypervariable regions are exposed and generally correspond with surface loops. Though the physiological significance of these changes is presently unknown, these hypervariable regions represent surface domains where amino acid substitutions could occur with little threat to the structural integrity of the molecules (Xie et al., 1991; Hughes et al., 2000). The buffalo mature peptide sequence also contains variable regions as in other species having profound nonsynonymous substitutions. There are mainly four variable regions at positions 101-117, 162-173, 241-248 and 331-350. Since these proteins belong to glycoprotein family, they possess distinct sites for glycosylation. The potential sites for glycosylation in deduced buffalo PAG-1 protein sequence are at positions 78-81, 127-129 and 250-300 (Guruprasad et al., 1996; Green et al., 1998) (Fig. 7).

Phylogenetic Analysis of the buffalo PAG-1 sequences with other domestic species revealed buffalo PAG-1 is more related with bovine, caprine and ovine species and less similar to equine, porcine, deer and zebra due to various nonsynonymous substitutions in the entire sequence in the latter species.

Porcine PAG-1 forms an entire different clade as it consists of variant amino acids residues when compared with ruminant species. Phylogram constructed between
buffalo PAG-1 and different bovine PAGs sequences revealed buffalo PAG-1 was more related with bovine PAG-1 thereby belonging to the recently duplicated PAG gene group (PAG-1 group) and differing significantly from the ancient PAG group consisting of PAG-2 and 8 (Vasques et al., 1995; Garbayo et al., 2000, Szafranska et al., 1995).

CONCLUSION
Buffalo Pregnancy Associated Glycoprotein-1 cDNA encoding buffalo PAG-1 protein of 380 amino acids was characterized and found to be recently duplicated group members of the aspartic proteinase family being proteolytically inactive due to key mutations close to the active site.

ACKNOWLEDGEMENTS
The researches are grateful to the Director Indian Veterinary Research Institute and Head, Animal Reproduction Division for providing the necessary facilities for conducting the experiment.

REFERENCES
Guruprasad, K., T.L. Blundell, S. Xie, J. Green and B. Szafranska et al., 1996. Comparative modeling and analysis of amino acid substitutions suggests that the family of pregnancy-associated glycoproteins includes both active an inactive aspartic proteinases. Protein Eng., 9: 849-856.

