



## Research article

Buffalo early pregnancy biomarker coding sequence cloning and partial length expression in *E. coli* after codon optimizationShree Vidhya Srinivasan<sup>a,b</sup>, Jyotirmoy Ghosh<sup>a,\*</sup>, Sayed Basha Nazar<sup>a</sup>, Kajal Sankar Roy<sup>c</sup><sup>a</sup> Molecular Biology Laboratory, Division of Physiology, ICAR-National Institute of Animal Nutrition and Physiology, Hosur Main Road, Adugodi-PO, Bangalore 560 030, Karnataka, India<sup>b</sup> Jain University, Bangalore, Karnataka, India<sup>c</sup> Division of Bioenergetics and Environmental Sciences, ICAR-National Institute of Animal Nutrition and Physiology, Hosur Main Road, Adugodi-PO, Bangalore 560 030, Karnataka, India

## ARTICLE INFO

## Keywords:

Veterinary medicine

Molecular biology

Buffalo

Pregnancy-associated glycoproteins

Recombinant PAG

## ABSTRACT

Pregnancy-associated glycoproteins (PAGs) secreted from conceptus specific trophoblast cells are widely accepted biomarkers of ruminants. Limited information of PAGs in buffalo warrants further investigation for the development of sensitive homologous early pregnancy-specific diagnostic immunoassay. This experiment was thus designed to identify and clone the predominantly expressed early placentome-specific buffalo PAG (buPAG) isoform; to express this PAG isoform and verify its antigenicity by developing antisera and testing immuno-reactivity with recombinant proteins. Results indicated PAG 7 (buPAG 7) was the predominant isoform in buffalo early pregnant placentome. Attempt to express the full native glycosylated protein in the pcDNA3.3 vector and FreeStyle HEK 293F host was not successful. However, a partial 124 amino acid sequence selected from the non-glycosylated region of buPAG 7 could be expressed in *E. coli* BL21 (DE3) cells after codon optimization however; the yield was low. Antigenicity of the expressed protein was confirmed by successful immuno-reaction in rabbits indicating possibilities of using 124 aa partial PAG 7 protein as a putative antigen for monoclonal antibody production and development sensitive homologous immunoassay. In conclusion, our results confirmed the findings that buPAG 7 as the predominant early pregnancy-specific transcript. A selected partial 124 amino acid sequences of it could even be expressed in a heterologous host (*E. coli*). Based on our data presented here, we anticipate that the expressed recombinant protein can be useful as an antigen suitable for the development of PAG specific immunoassays in buffalo.

## 1. Introduction

Pregnancy-associated glycoproteins (PAGs) are multi-isoform proteins generated by gene duplication probably due to continuous selection pressure at the maternal-conceptus junction [1]. They are products of mono- and bi-nucleated trophoblast cells, identified in cattle, buffalo, swine, camel, wild deer and bison placenta [2], released extracellularly by exocytosis thus available in maternal circulation [3] and used for immunodetection of early pregnancy in ruminants. About 100 PAG genes in bovine [4] and 51 coding PAG transcripts in other ruminant species [5, 6] are reported in the literature. Variation in isoforms, expression patterns across the species and pregnancy stages [4] along with sequence similarities with serum proteases add complexities, and challenges of species-specific immunodetection of this protein. Because of these

reasons, PAG detection is likely to be less sensitive in heterologous and common ruminant assays. A specific, sensitive and accurate immunoassay demands a homologous antigen for raising 100 % cross-reactive antibodies. In buffaloes, about 22 PAG mRNA sequences are available in NCBI/UniProtKB/TREMBL database however; limited information on expression patterns of different isoforms during pregnancy stages warrants further investigations. This information would be critical for the development of buffalo specific early pregnancy diagnosis assay. The PAG of buffalo has been purified by conventional chromatography and used as the standard for heterologous radio-immunoassay design and also for profiling month-wise secretion in maternal blood during pregnancy [7]. *In silico* 3D structures are predicted from deduced amino acid sequences of buffalo PAG1 [8], PAG2 [9], PAG7 [10,11] and other isoforms [11] to understand their structural properties and functional roles.

\* Corresponding author.

E-mail address: [jghosh68@gmail.com](mailto:jghosh68@gmail.com) (J. Ghosh).

Available literature so far indicated that PAG7 is the predominant early pregnancy-specific isoform in buffalo placenta [11] however; the full coding sequence is not reported.

Purification of early pregnancy-specific PAG isoform through conventional chromatography approach is not feasible as purification of the desired isoform might require a huge quantity of starting tissue [7]. Multiple fresh early pregnant placentas can only be obtained by the sacrifice of the pregnant animal, not permissible on ethical and technical ground. It is also likely that when placentas are pooled from multiple sources, the desired isoform does not get purified from the source tissue, due to different biochemical properties. On the other hand, specific isoform transcript can be identified from milligram quantities of tissues; cloned and recombinant protein produced in suitable host cells.

Use of a native glycosylated PAG as an antigen is of paramount importance for raising specific antibody. It is because MHC (more precisely class II) molecules in antigen presenting cells would present the antigen naively for T-cell recognition [12] and therefore, any alteration of native protein structure would result in production of antibody with lesser reactivity and lower specificity. *In silico* epitope prediction revealed that usually the glycosylated areas of a protein are hydrophilic and contain immunodominant epitopes [13]. However, antigenic epitopes may also be present in the non-glycosylated regions of a protein and they can serve as targets for antibody production [14]. Thus, experiment was designed to identify the predominant buffalo PAG isoforms coding sequences of all the isoform sequences. The ORFs (open reading frames) of these isoforms were cloned and sequenced. The partial sequence of the most predominant PCR product was expressed in *E. coli*, after codon optimization. The recombinant protein was used for raising antisera, and antigenicity was tested by immunoreactivity of recombinant protein against its antisera.

## 2. Materials and methods

### 2.1. Chemicals

All the chemicals used for the experiments were of analytical grade and purchased either from Sisco Research Laboratory (SRL) or HiMedia Laboratories, Mumbai, India if the sources were not disclosed.

### 2.2. Bio-safety clearance

All methods for this study were carried out with necessary approval and following the relevant guidelines and regulations of the Department of Biotechnology, New Dehli and ICAR-National Institute of Animal Nutrition and Physiology, Bangalore, India.

### 2.3. Collection of tissues

The whole buffalo reproductive tract from the ovary to the cervix was collected from a local abattoir (Corporation Slaughter House of Bangalore, Karnataka, India) immediately after slaughter and brought to the laboratory on ice. The appearance of uterine horns (symmetry/asymmetry) was confirmed visually and by hand palpation. The corpus luteum in ovaries was examined; isolated and morphological characteristics were assessed as per the published method [15]. Horns of the uterus were cut open using scissors and observed for the presence and absence of conceptus (fetus and placenta), pathological lesions and structures. Pregnancy was confirmed by the presence of conceptus. The whole fetus was isolated from the conceptus mass and crown-rump length was measured for an approximate determination of the day of pregnancy using the formula of Soliman et al [16]. Tissue samples were harvested at luteal phase, identified by the presence of corpus luteum on ovarian surface [15]. Necessary precautions were taken to ensure samples from reproductive track with symmetrical uterine horns free from pathological lesions. About 100 mg tissues from luteal phase non-pregnant endometrium (n = 3), caruncle from pregnant animals (approx. Day 55–67) (n =

3) and placental cotyledon (n = 3) were collected and stored in RNA-later at 4 °C overnight and then at – 80 °C for long term storage.

### 2.4. Isolation of total RNA and cDNA synthesis

The total RNA from 50 mg tissues of each sample was isolated using the RNeasy Plus Mini kit (Qiagen, Germany) with a DNA removal step by in-column DNase digestion with RNase-Free DNase Set (Qiagen, Germany). Quantity of total RNA was estimated by NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific™, USA) and the quality was tested with 1 µg of total RNA resolved in Agarose (1%) gel by electrophoresis in 1X MOPS [3-(N-morpholino) propane-sulfonic acid] buffer.

Single-stranded cDNA was synthesized from 2 µg of total RNA of each sample using Revert Aid™ H minus first-strand cDNA synthesis kit (Thermo Fisher Scientific, USA) following manufacturer's instructions.

### 2.5. PCR amplification, cloning, and sequencing of predominant buffalo PAG

All cDNA samples synthesized from RNA isolated from caruncle (of both pregnant and non-pregnant animals) and cotyledon (of pregnant animals) were screened for the presence of open reading frames of all buffalo PAG reported previously. This was done by PCR amplification in 20 µL reaction volume with a fixed single forward primer (L) and each of four different reverse primers R1–4 [Supplementary Table 1]. Each reaction mixture contained 20 ng cDNA templates, 2X master mix, and 5 µM each of forward and reverse primers (Promega, USA). The amplification conditions were set as one cycle of denaturation at 94 °C for 2 min, 35 cycles of denaturation (94 °C for 30 s), annealing (55 °C for 30 s) and extension (72 °C for 90 s) and final one cycle of extension at 72 °C for 5 min. Once the predominant PCR product was identified, PCR was repeated using proofreading Accuzyme DNA polymerase (Bioline, USA, 2.5 U/µL) and the same primer set. The intense PCR bands from luteal phase non-pregnant caruncle and day 55 pregnant cotyledon and caruncle samples were gel purified using illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, USA) and used for ligation in pJET1.2/blunt cloning vector. The One Shot® TOP10F *E. coli* competent cells (Thermo Fisher Scientific Inc, USA) were used for transformation with the ligation mixture. The clones were selected in the presence of Ampicillin in plates and broths. The cloned plasmids were isolated by Miniprep Kit (PLN70, Sigma Aldrich, USA) and the inserts were confirmed by PCR using specific primer set [see Supplementary Table 1] and products were sequenced from both directions.

### 2.6. Authentication and analysis of the sequence

Sequence data was analyzed by NEB cutter (<http://nc2.neb.com/NEBcutter2/>) for identification of suitable restriction enzymes and expected fragment sizes, if digested with those enzymes. Accordingly, plasmid was digested with different restriction enzymes (New England Biolabs, UK) in 20 µL reaction volume viz., BamH I (R0136S), Bgl II (R0144S), Eag I (R0505S), PmeI (R0560S), Sal I (R0138S) and Xho I (R0146S). Based on electrophoresis, sizes of fragments were determined and compared with the predicted fragment sizes generated with NEB cutter. The authenticated sequence obtained by this method was annotated by the NCBI nucleotide BLAST tool ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=blastSearch](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=blastSearch)); phylogenetic similarity was determined by MEGA 7 [17]; coding and translated amino acid sequences were obtained by NCBI tool (<https://www.ncbi.nlm.nih.gov/orffinder/>). Screening for the presence of probable 20 amino acid long linear epitopes was done by BCPREDS online tools (<http://ailab.ist.psu.edu/bcpred/predict.html>).

### 2.7. Expression of buffalo PAG in mammalian host cells

The full coding sequence was amplified by PCR from the pJET1.2-

buPAG7 vector and integrated into the pcDNA3.3 mammalian cell expression vector. Top10 *E. coli* competent bacteria available with Invitrogen™ pcDNA™3.3-TOPO™ TA Cloning™ Kit were transformed with the resulting recombinant pcDNA™3.3-buPAG7 vector. The plasmid was isolated by Mini-prep and sequenced from both directions to confirm the orientation and reading frame of the expression cassette. The confirmed expression cassette was produced in bulk and purified using PureLink® HiPure Plasmid Filter Midiprep Kit (M/s Thermo Fisher Scientific, USA). One T 175 flask of FreeStyle™ HEK 293-F (M/s Thermo Fisher Scientific, USA) cells were grown in suspensions and transfected with either the test vector or with a positive control, a construct encoding lacZ cassette (M/s Thermo Fisher Scientific, USA). For transfection, 293fectin™ Transfection Reagent (M/s Thermo Fisher Scientific, USA) was used. After 48 h of transfection, the cells and the spent media were harvested for analysis by Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Attempts to express the protein were repeated 3 times but failed to observe the specific band each time (see the results).

## 2.8. Codon optimization of partial buPAG by PCR primer based approach

A 372 bp sequence (495–866 bp) was selected with an ATG at the beginning in the cloned PAG sequence of buffalo for expression. Attempts were made to modify 7 rare codons (AGA to CGT, CTA to CTG, CCC to CCG, CGC to CGT, AGA to AGT, AGA to CGT, ATA to ATC) using two sets of PCR primers [Supplementary Table 2]. Initially, two 50 µL PCR reactions were carried out using two different primer sets, each with 50 ng template (pJET1.2-buPAG), suitable buffer and proofreading Accuzyme DNA polymerase (2.5 U/µL; Sigma Aldrich, USA). The PCR was performed at initial 95 °C denaturation for 5 min and 35 cycles of denaturation at 95 °C for 5 s, annealing at 55 °C for 10 s and extension at 72 °C for 10 s. The PCR products were separated by 1.5 % Agarose gel, bands were excised from the gel, and purified as described earlier. Purified PCR products from both reactions were used as templates for the second round of PCR using a forward primer of set 1 (PP1-F) and the reverse primer of set 2 (PP2-R). The cycling conditions were 95 °C denaturation for 5 min followed by 35 cycles of denaturation at 95 °C for 10 s, annealing at 55 °C for 10 s and extension at 72 °C for 10s. The final PCR products were resolved, purified from gel as explained previously and sequenced using gene-specific primers to confirm the base changes.

## 2.9. Sub-cloning of PCR product in pET 28a (+) vector

Upon confirmation of base changes, the PCR product was inserted into the pET-28a(+) vector (# 69864-3, Novagen, USA). About 1 µg of pET-28a (+) vector and 2 µg of PCR product were digested using Nde I and Xho I (New England Biolabs Inc. UK) restriction enzymes (RE) at 37 °C for 2 h, separated in 1.5 % agarose gel by electrophoresis, specific bands were excised and purified. Gel purified 50 ng vector and 19 ng insert (1:5) were ligated at 16 °C overnight using T4 DNA ligase (# M0202S, New England Biolabs Inc. UK) and the newly generated recombinant vector was designated as pET-28a(+)-par.buPAG. Recombinant pET-28a (+)-par.buPAG vector was used for the transformation of One Shot® TOP10F chemically competent *E. coli* bacteria and selected by the supplementation of antibiotics Kanamycin (Kanamac®-1000, Macleods Pharmaceuticals Ltd, India) in culture media. Plasmids were isolated from the antibiotic selected clones, insert was confirmed by PCR. The 3 selected clones were sequenced by Sanger's chain termination method using vector-specific sequencing primers to confirm the reading frame and coding sequences.

## 2.10. Expression of recombinant partial buffalo PAG in *E. coli* and purification

The BL21 (DE3) competent *E. coli* cells (# 200131, Agilent Technologies, USA) were transformed with recombinant pET-28a (+) par.buPAG vector having proper orientation and reading frame. Clones were

selected by growing them in the presence of Chloramphenicol (Calbiochem, EMD Biosciences, Germany) and Kanamycin antibiotics. The protein production in *E. coli* cells was induced by 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Calbiochem, EMD, Biosciences, Germany). A similar culture without IPTG induction was prepared for testing as control. Expression of proteins in cells and media was confirmed initially by 15 % SDS-PAGE analysis gel stained with Coomassie® Brilliant blue R-250 dye. Upon confirmation from cells, protein was harvested in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 % glycerol) containing cOmplete™ Protease Inhibitor Cocktail (Roche Applied Science, Germany). Bulk quantity of proteins was harvested using Ni-NTA Superflow Cartridge (# 30721, Qiagen, Germany). In brief, the cell pellet was sonicated in NPI-10 buffer [50 mM Sodium phosphate monobasic (NaH<sub>2</sub>PO<sub>4</sub>), 300 mM Sodium chloride, 10 mM Imidazole (Sigma Aldrich now Merck-Millipore, USA), pH 8], cell lysate was clarified by centrifugation at 10,000 g for 30 min at 4 °C, then the supernatant was passed 3 times through the Ni-NTA superflow cartridge. The final unbound protein was removed by thorough washing with 50 mM NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 8.0) containing 300 mM NaCl, 20 mM Imidazole and the bound protein was eluted in the same buffer with higher amount (250 mM, in place of 20 mM) of Imidazole. Small volumes from the eluted fractions were tested for the presence of expressed protein in 15 % SDS-PAGE. The eluted bound protein fractions from different columns were pooled and purified again by passing through the Ni-NTA column and dialyzed against 1 X PBS and quantified by Lowry's protein estimation method [18].

## 2.11. Development of polyclonal antibody and purification

The polyclonal antibody was raised in two rabbits by immunization with purified recombinant partial buPAG proteins following the standard protocol of one primary and three booster injections. The generated polyclonal antibody was purified using protein-A Agarose column (Pierce™ Protein-A Column, # 20356, Thermo Fisher Scientific, USA). The antibody present in sera was allowed to bind with the column matrix in the presence of binding buffer (PBS: 0.1 M sodium phosphate, 0.15 M sodium chloride, pH 7.5). The unbound flow-through and the column wash were discarded. The bound antibodies (IgG) were eluted in 0.1 M glycine buffer (pH 3) and fractions were collected in 10 tubes with 0.5 mL each. The pH of each tube was raised immediately by adding 50 µL 1 M phosphate buffer (pH 8.5). Absorbance at 280 nm of each tube contents was measured and fractions showing the positive absorbance against the buffer were pooled and dialyzed against PBS, pH 7.5 by using 14 kDa molecular weight cut-off bags (# D9652, Sigma Aldrich, USA).

## 2.12. Western blot analysis of recombinant protein

About 100 µg of cell extracted proteins from each group (induced and non-induced *E. coli*) were separated by SDS-PAGE and transferred onto PVDF membrane (Immobilon® 0.45 µM, # IPVH00010, Merck-Millipore, USA) for Western blotting. The PVDF membrane was sandwiched with the SDS-PAGE gel and the transfer of proteins was carried out in presence of 1X blot transfer buffer [25mM Tris buffer, 192 mM Glycine, 10% methanol] at 50 V for an initial 2 h followed by at 30 V overnight. The non-specific protein bands on the membrane were blocked by incubating the membrane overnight at 4 °C in 3 % BSA dissolved in 1X PBS. This was followed by incubation of the membrane with 1:5000 dilutions (1X PBS) of primary antibody (Rabbit anti-PAG) for 1 h at room temperature followed by 3 washing with sufficient quantities of 1X PBS/Tween 20 (PBST). The membrane was then incubated with 1:1000 diluted goat anti-Rabbit-HRP conjugated secondary antibodies (Thermo Fisher Scientific, USA) for 1 h at room temperature. The unused secondary antibody on the membrane was washed with PBST three times and finally, protein bands that reacted to primary as well as secondary antibody were developed using SIGMA FAST™ DAB with Metal Enhancer (D0426, Sigma Chemicals) following manufacturer's protocol.

### 3. Results

#### 3.1. Early pregnancy specific predominant PAG transcript in placenta

In the current experiment, buffalo PAG transcripts were amplified by PCR reaction and resolved by gel electrophoresis. Band movement patterns of products amplified from caruncle and cotyledon tissue of early pregnant buffaloes were similar. A very intense band was observed between 1 kb to 1.5 kb region in addition to many low-intensity bands below 1 kb by L-R1 pair (Fig. 1A). The L-R2 pair produced many low-intensity bands between 0.6 to 0.3 kb, L-R3 pair did not show any band, and L-R4 pair produced faint bands near 1.5 kb, 0.5 kb, and 0.3 kb region. The non-pregnant caruncle showed an intense band near 0.6 kb (Fig. 1A). A major PAG specific cotyledon band was identified to have 1163 bp nucleotides in the three tested clones. The products from the restriction enzyme digestions were in agreement with the results from the NEB cutter (Fig. 1D). BLAST analysis revealed it had 99 % sequence identity to buffalo PAG 7 (HM\_468476) and the alignment covered 95 % of the previously published buffalo PAG 7. Cloned buffalo PAG7 obtained in the present experiment had 97 % sequence identity to bovine PAG7 (NM\_001109978) at 99 % coverage. The sequence has been deposited in GenBank (KP233844) (<https://www.ncbi.nlm.nih.gov/nucleotide>). Phylogenetic analysis revealed the sequence was closest to cattle followed by bison, sheep, goat, zebra, donkey, and horses [Supplementary Fig. 1]. Besides it had 98 % and 93 % similarity to buffalo PAG11 (HM\_468480.1) and PAG1 (XR\_329125.1), respectively but had lower similarities with other buffalo PAG isoform sequences. The sequence was deduced to code for protein having 380 amino acids (aa) (from 1140 bp) with a theoretical molecular weight of about 42.98 kDa [Supplementary Table 3 (A) & (B)]. Initial 15 aa contained secretory signal and the next 28 aa (from 16-43) was propeptide. It also contained pepsin- A (cd05478,

62-377 aa), eukaryotic aspartyl protease (Asp, pfam0026, 70-379 aa) pepsins, cathepsins, and renins, PTZ00165 aspartyl protease (63-379 aa) and A1-Propeptide aspartic protease (pfam07966, 16-43 aa domains). At the amino acid level, the sequence was 96 % (356 of 370 aa at the matched region) similar to PAG 11 [Supplementary Table 4].

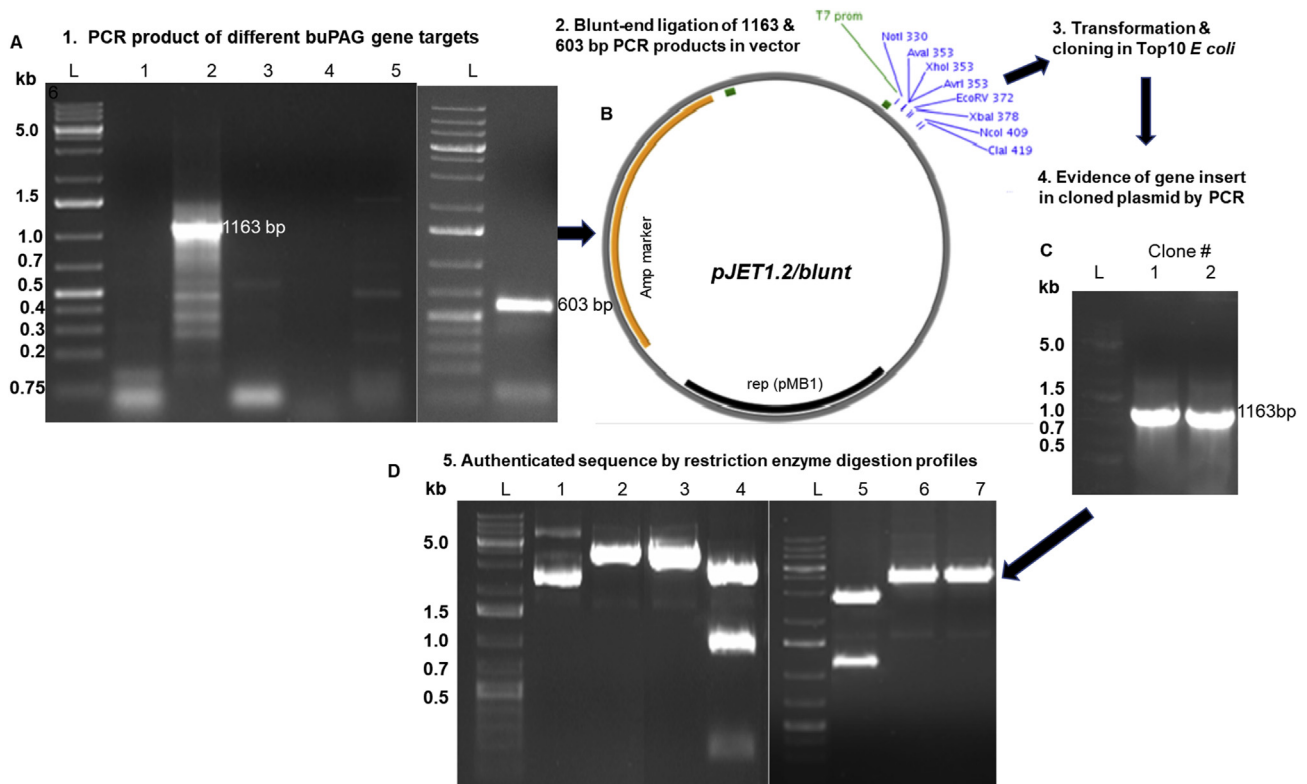
The polymerase chain reaction amplified 603 bp transcript of PAG gene from caruncle of non-pregnant buffalo. The first 108 bp of this shorter transcript was identical to 5' sequences of transcript isolated from cotyledon of pregnant buffalo (KP233844) and the remaining 495 bp also matched to 3' end PAG7 sequences (KP233844). The 603 bp transcript sequence had a start codon (ATG) and a stop codon (TAA) at the similar location of 1163 sequence (KP233844). In addition, 603 bp transcript has multiple intervening TGA stop codons (data not shown), possibly terminating translation prematurely.

#### 3.2. Expression full length buffalo PAG major transcript in mammalian cells

The SDS-PAGE analysis revealed that the positive control cells expressed the LacZ protein-specific protein band but not the test cells and its media supernatant (data are not shown). In all the three attempts, appearance of the expected 40–84 kDa band for buPAG expression was not observed on SDS-PAGE analysis gels.

#### 3.3. Expression of partial buffalo PAG in *E. coli* cells

Selected 124aa sequence had 7 rare codons (<http://people.mbi.ucla.edu/sumchan/caltor.html>), no glycosylation sites (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and three predicted linear epitopes (<http://ailab.ist.psu.edu/bcpred/predict.html>). The results of PCR and sequence analysis indicated a successful mutation (Fig. 2B), a correct



**Fig. 1.** (A) Amplification of PAG in buffalo cotyledon and non-pregnant endometrium samples using single L and four right (R1-R4) primer pairs. Lanes: 1-RNA, 2-L-R1, 3-L-R2, 4-L-R3, 5-L-R4, L- DNA ladder (Gene Ruler™ 1kb DNA Ladder Plus cat no N3232S, NEB, USA); (B) Cloning of 1163 and 603bp PCR products in pJET1.2 vector; (C) Amplification of isolated plasmids by gene-specific primers; (D) restriction enzymes digestion profiles of positive plasmids (L: DNA ladder, lane 1: uncut cloned vector, 2: BamH I digest (single cut), 3: Sal I digest (single cut), 4: Xho I digest (3 cut sites:1164bp, 990bp and 170bp products), 5: Bgl II digest (3 cut sites-1194bp, 1069bp and 113bp products), 6: Eag I digest (single cut) and 7: Pme I digest (single cut) indicating the authenticity of sequence reads.

orientation and correct reading frame of the expression cassette. All the seven codons were found mutated (R: AGA to CGT, L: CTA to CTG, P: CCC to CCG, D: CGC to CGT, K: AGA to AGT, G: AGA to CGT, I: ATA to ATC) However, while optimizing codon two un-deliberate base change was introduced at position 26 (G: GGC to V: GTC) and at 243 (W: AGC to AGT) of 372 bp sequences [Supplementary Table 5]. This resulted in change of 9 codons but substitution of only one amino acid (from G to V) and right shift of the first epitope keeping the core 16 aa sequence (ILGLNYRNLSWSKTLTP) same. However, position of the other two epitopes remained unchanged [Supplementary Table 6]. Proteins isolated from *E. coli* cell expressing partial buffalo PAG sequence showed to have an 18 kDa (Fig. 2C) intense band in the gel. The Ni-NTA column purification resulted in an enrichment of 18 kDa band in the final purified product (Fig. 2D).

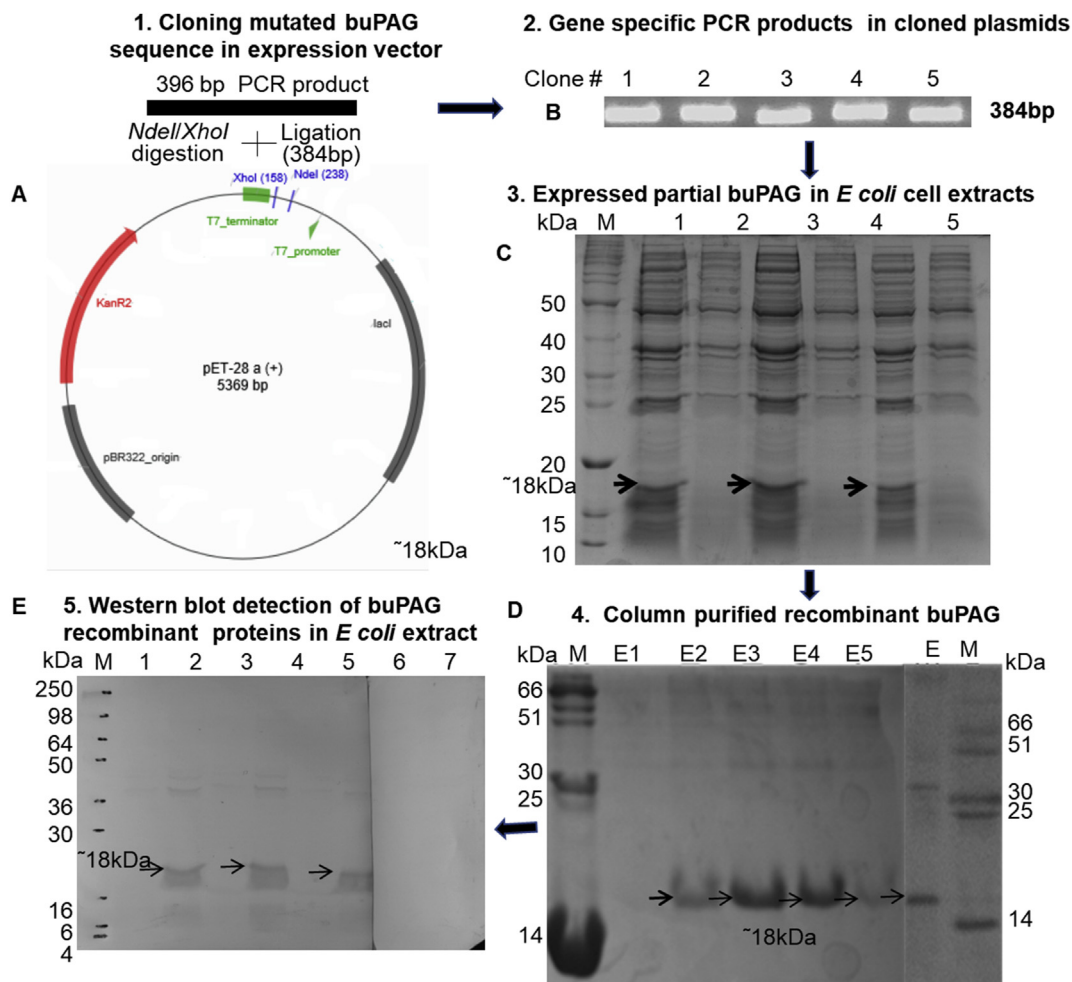
### 3.4. Immunoreactivity of purified polyclonal antibody with recombinant partial buffalo PAG

Antisera production in rabbits was successful against partial buPAG expressed in heterologous host, *E. coli*. Western blot analysis revealed affinity purified antibody reacted with two adjacent prominent bands near 18 kDa region of the whole protein extracts expressed in IPTG induced *E. coli* of but not in the extracts of non-induced control bacteria (Fig. 2E).

## 4. Discussion

Identification of early pregnancy-specific biomarkers is of immense importance and an urgent prerequisite for the development of a sensitive early-pregnancy diagnosis assay in buffaloes. This study reports full open reading frame sequence information of a predominant buffalo early pregnancy-specific PAG isoform, recombinant expression of partial sequence (124 aa) in *E. coli* after codon optimization, the antigenic potential of the recombinant protein by successful antibody production and immunoreactivity.

The nucleic acid (KP233844) and deduced protein (AKB95131.1) sequences is named as PAG 7 sequences. This was based on very high percent of identity of sequence with published buffalo (HM\_468476.1) and of other species PAG 7 isoforms. The authentication of the sequence was confirmed by matching of experimental and theoretical (restriction enzyme digestion profiles) results. The presence of signal peptide, pro-peptides, mature protein and different domains in the derived sequences were as per the published reports on PAG of other species. Therefore, derived sequence was considered as the full reading frame of buffalo PAG 7. Similar to our results, PAG 7 has recently been reported as the predominant early pregnancy-specific isoform in buffalo using a different approach where three different forward and single right primers were used for the study [11]. In our approach identification of the predominant isoform was much easier and less expensive than the other report since we only cloned logically the most intense PCR product to



**Fig. 2.** Sub-cloning and expression of 372bp mutated sequences (Extra sequences indicated in the label due to RE sites on both sides of primers). A) Schematics of cloning approach; B) Clones showing desired PCR amplification; (D) Expressed protein bands in 15 % SDS-PAGE gel image in different column eluents (E1-E5) and a pooled fraction (E). E) Western blot image showing PAG immunoreactive protein bands in lanes 2, 4, 6). Lane M - protein marker (Benchmark TM protein ladder, Thermo Scientific, USA). Lanes: 1, 3, 5 - IPTG non induced cell extract; 2, 4, 6 - induced cell extract; 7, 8- are a negative test (no primary antibody).

arrive at the decision. Additionally, we amplified and cloned the full open reading frame of the particular isoform. Studies of others in buffaloes indicated that early pregnant cotyledon also expresses low abundant buPAG 18, 2, 8, 6, 9, 13 and 15 isoforms [11]. The low abundant PCR products (Fig. 1) proved the presence of other isoforms. However, identities of those isoforms were not confirmed by cloning and sequencing in our study since we focussed only on the sequence and identity of the predominant isoform. The information generated in our study and others [11] might help for further downstream applications and development of sensitive immunoassay for buffalo early pregnancy biomarker detection. As compared to our observation in buffalo, PAG 2 was found to be the predominant isoform in cattle [19]. A pregnancy stage-dependent expression of PAG isoforms were reported in cattle and other species. Binucleate cells of cattle although predominantly express PAG 1, 6, and 7 isoforms in middle and late pregnancy, the day 25 and earlier cotyledons either do not or weakly express these isoforms. On the contrary, cotyledons express PAG 4, 5 and 9 isoforms in abundance at early stage (25 days or earlier) of pregnancy [3]. A cDNA microarray analysis in cattle demonstrated expression of several PAG molecules between day 7–14 of pregnancy (PAG 11, 16 and 17), day 14–21 (PAG 1, 5 to 7, 9 to 13, 15 to 17, 19, 21) or even before (at day 7: PAG 4, 5 and 6) [20]. Similar cDNA based microarray study in buffalo revealed the PAG 4–8, 11 and 17 were the top up-regulated isoforms in fetal cotyledons during early pregnancy [21]. In equine PAG 1 and porcine PAG 1 and 2 were found as the predominant isoforms expressed in the chorion [19, 20].

The PAGs were purified from the placenta of cattle [22], sheep [23], goat [24], pig [25], horses [26] and buffaloes [7]. The antibodies produced against PAG isoforms helped the development of immunoassays and pregnancy diagnostics in different species. However, early pregnancy diagnosis in buffaloes using the existing heterologous and common ruminant assays is not successful. The mismatch of sequences and less specific immunoreactivity of antibodies with the early pregnancy-specific PAG isoforms of buffaloes might have contributed to this. The diagnosis of early pregnancy is thwarted without the use of specific and best reactive monoclonal antibodies against early pregnancy-specific buffalo PAG.

In placenta, each isoform of PAG is usually expressed in a very low quantity and therefore, to purify sufficient amount of any of these isoforms, one needs to have good number of placenta from buffaloes at early stage of their pregnancy. If PAG is purified from placenta of multiple sources, heterogeneity in samples may pose difficulty in further use. In addition, most isoforms of 22 different PAGs have shared properties and therefore, it poses technical difficulty to purify a specific isoform. Moreover, in most societies, slaughter of huge number of buffaloes for collecting placenta at early pregnancy is not ethically permissible. Due to these reasons, the productions of desired PAG by recombinant technology are preferred over conventional purification. On the contrary, designing primers to amplify the specific isoform with the help of reported sequences, cloning and sub-cloning for expression in suitable cells are easy although, selection of vector, host, and standardization of expression protocols need time and effort. For expressing cattle PAG, Green monkey kidney (COS) and Chinese hamster ovarian (CHO) cells [27] were not found to be good hosts due to lower yield of recombinant PAG as compared to HEK 293F host cells [28]. Therefore, we used HEK 293F cells for expression of full length buffalo PAG protein. We preferred the expression of full length sequence because it contained 9 predicted antigenic epitopes. In this experiment, the full length nucleotide sequence of buffalo PAG 7 was cloned under CMV promoter in pcDNA 3.3 mammalian expression system. This mammalian expression system was reported to provide high transient expression of recombinant proteins [29].

However, our attempt to express the protein was not successful probably due to the presence of rare codons in the sequence. Also in the mammalian host cells, it might have yielded truncated or low amount of the protein. PAG is generally produced by trophoblast cells whereas HEK

293F cells were of kidney origin, therefore, unsuitable for sustained expression of transgene. Poor transfection efficiency of transgene in to host cells could also be responsible for low of transgene expression. These issues can be addressed with appropriate strategies.

Analysis of 380 aa buffalo PAG 7 sequence revealed that out of 9 predicted epitopes, 4 were present in the predicted N-glycosylation sites and 5 in the non-glycosylated regions [Supplementary Table 6]. Shreds of evidence suggest that antibodies produced against a non-glycosylated region of a protein had specific reactivity against the native glycosylated proteins [30, 31, 32]. In order to take advantages of availability of predicted epitopes in non-glycosylated region, a stretch of 372 bp (124 amino acids) sequences that covered three predicted epitopes were selected for expression in *E. coli* system. We considered the partial sequence of PAG rather than full length of it for expression in *E. coli*. The reason for such consideration was that *E. coli* based expression system cannot support formation of native structure and proper folding of the full protein due to the absence of enzymes needed for glycosylation and other post-translational modifications. In the selected 124 aa sequence, 3 predicted antigenic epitopes were present but not any glycosylation site [Supplementary Table 6]. The nucleotide sequences of the selected stretch were 100 % identical to buffalo PAG 7 and PAG 11 (Supplementary Table 4). This similarity in the epitope regions may have certain advantages, if used for raising antibody for detecting PAG 7, PAG 11 and other similar isoforms if expressed during early pregnancy.

Proteins are often difficult to express outside their native context, particularly if they contain rare and non-canonical codons, not supported by the host. Sometimes, expression-limiting regulatory elements within their coding sequence may also prevent expression of proteins [33]. Thus for expression of heterologous protein in a prokaryotic system, codon usage has been identified as the single most important factor [34]. The preferred codons correlate with the abundance of cognate tRNAs available within the cell that in turn serves to optimize the translational system and to balance codon concentration with iso-acceptor tRNA concentration [35]. Thus, a gene with more rare codons is less likely to be expressed at reasonably good levels in the host [36]. A common strategy to improve expression is thus to alter rare codons of the target gene that closely reflect the codon usage pattern of the host cells, without modifying the amino acid sequence of the encoded protein. A common technique to achieve this goal is to cause sequential site-directed mutagenesis by introducing desired changes in PCR primers. We have adopted this strategy to alter the rare codons present in the selected 372 bp sequence for expression in *E. coli*. The expression of mammalian proteins in the *E. coli* host has been demonstrated by this approach in previously published report [37]. In general optimization of codon increases expression dramatically in *E. coli* host in some cases [38] but in some instances, proteins were just expressed successfully, otherwise could not be expressed with the native sequence [39, 40]. The codon optimization might increase the expression of mammalian proteins typically by 5–15 folds in the *E. coli* system and yields might reach as much as 5 % of their soluble proteins [33]. Our approach of codon optimization helped the production of 13 mg protein from 2 L *E. coli* culture pellets 48 h after IPTG induction. The amount of protein expressed was sufficient for immunization of two rabbits (10 mg), for SDS-PAGE and Western blot analysis (1 mg) and stored for future downstream applications (2 mg).

Heterologous antibodies might cross-react with buffalo PAG [41] however; antibodies for a wrong isoform may not help the detection of early pregnancy signals because of the lower percentage of cross-reactivity and segregation of non-pregnant and early post-partum animals. Experiment in cattle revealed that, PAG1 is not an ideal pregnancy biomarker due to their expression between ~ day 45 up to 250 of pregnancy and higher expression at term and availability during post-partum [3]. The reason for availability of untranslated prominent short 603 bp PAG-like transcripts in the non-pregnant buffalo caruncle is not clear.

Identification of PAG 7 as the predominant transcript, recombinant expression and successful antibody production opened up a new avenue

of research on PAG in buffalo. Successful polyclonal antibody production in rabbits indicated the antigenic potential of the recombinant protein. The purified antibody found cross-reacted with the recombinant buffalo PAG confirmed that 124 aa recombinant protein was antigenic and can be used for development of monoclonal antibody, a prerequisite for development of a sensitive homologous PAG assay kit for early pregnancy diagnosis. Considering this sequence for epitope regions is 100 % similar to PAG 7, PAG 11 and other isoform epitopes, sequences PAG polyclonal antibodies produced under the present study should be capable of capturing signals from multiple isoforms.

In conclusion, we report here the authentic full open reading frame sequence of the predominant buffalo early pregnancy-specific PAG 7 isoform, *in silico* properties, successful expression of 124 aa sequences in *E. coli* after codon optimization. We also demonstrate the antigenic potential of the recombinant protein showing its immuno-reactivity with the raised antibody. Our study thus puts a step forward toward establishing PAG 7 as a functional early pregnancy biomarker in buffaloes and gives a clue that the recombinant 124 aa partial buffalo PAG protein can serve as an antigen for the production of monoclonal antibodies for further development of buffalo early pregnancy immunodiagnostic method in future.

## Declarations

### Author contribution statement

Shree Vidhya Srinivasan: Performed the experiments; Contributed in analysis and interpretation of data; Wrote the paper.

Jyotirmoy Ghosh: Conceived, designed and performed the experiments; Analyzed and interpreted the data; contributed reagents, materials, analysis tools or data; Wrote the paper.

Nazar Syed Basha: Contributed materials and analysis tools.

Kajal Sankar Roy: Contributed performing experiment.

### Funding statement

This work was supported by the Department of Biotechnology, Ministry of Science and Technology, New Delhi, India. BT/PR14884/AAQ/01/461/2010 and ICAR-NIANP, IRC APR-3.12.

### Competing interest statement

The authors declare no conflict of interest.

### Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2019.e02863>.

## Acknowledgements

Director, ICAR- NIANP for providing laboratory facilities, Jain University for allowing Ph.D. of the first author, M/s Chromus Biotech, Bangalore for *E. coli* based protein expression services, Dr. Sujoy K.Dhara, Stem cell Laboratory, Veterinary Biotechnology Division, Indian Veterinary Research Institute, Bareilly for correcting language for the entire manuscript.

## References

- [1] J.A. Green, Defining the function of a prolactin gene family member, *Proc. Natl. Acad. Sci. USA* 101 (2004) 16397–16398.
- [2] J.F. Beckers, P.V. Drion, J.M. Garbayo, Z. Perényi, A. Zarrouk, J. Sulon, B. Remy, O. Szenci, Pregnancy-associated glycoproteins in ruminants: inactive members of the aspartic proteinase family, *Acta Vet. Hung.* 47 (1999) 461–469.
- [3] J.A. Green, S. Xie, X. Quan, B. Bao, X. Gan, N. Mathialagan, J.F. Beckers, R.M. Roberts, Pregnancy-associated bovine and ovine glycoproteins exhibit spatially and temporally distinct expression patterns during pregnancy, *Biol. Reprod.* 62 (2000) 1624–1631.
- [4] S. Xie, J. Green, J.B. Bixby, B. Szafranska, J.C. DeMartini, S. Hecht, R.M. Roberts, The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes, *Proc. Natl. Acad. Sci. USA* 94 (1997) 12809–12816.
- [5] J.M. Garbayo, B. Serrano, F. Lopez-Gatius, Identification of novel pregnancy-associated glycoproteins (PAG) expressed by the peri-implantation conceptus of domestic ruminants, *Anim. Reprod. Sci.* 103 (2008) 120–134.
- [6] G.A. Vawter, T.E. Parks, J.A. Green, A.D. Ealy, G. Killian, Nucleotide Sequence for Pregnancy-Associated Glycoproteins From White Tailed Deer (*Odocoileus virginianus*), 2004. Direct submission to Gen- Bank [Acc nos Q6R6N3 to Q6R6N9 and Q6R6P0 to Q6R6P2].
- [7] O. Barbato, N.M. de Sousa, V.L. Barile, C. Canali, J.F. Beckers, Purification of pregnancy-associated glycoproteins from late-pregnancy *Bubalus bubalis* placentas and development of a radioimmunoassay for pregnancy diagnosis in water buffalo females, *BMC Vety. Res.* 9 (2013) 89.
- [8] A. Jerome, S.K. Singh, S.K. Agarwal, M. Saini, A. Raut, Characterization and *in silico* analysis of pregnancy-associated glycoprotein-1 gene of buffalo (*Bubalus bubalis*), *Genet. Res. Intern.* 43 (2011) 6138.
- [9] B. Ganguly, S. Prasad, Homology modeling and functional annotation of bubaline pregnancy-associated glycoprotein 2, *J. Anim. Sci. Biotechnol.* 3 (2012) 13.
- [10] S. Shree Vidhya, J. Ghosh, Homology modeling of PAG 7 and comparative interaction analysis of different PAG isoforms in buffalo (*Bubalus bubalis*), *J. Vety. Sci. Technol.* 6 (2015) 60.
- [11] M. Lotfan, S. Choudhary, M.L. Yadav, S. Kumar, S. Singh, S. Bathla, et al., Primary structures of different isoforms of buffalo pregnancy-associated glycoproteins (BuPAGs) during early pregnancy and elucidation of the 3-dimensional structure of the most abundant isoform BuPAG 7, *PLoS One* 13 (2018) e0206143.
- [12] M.A. Wolfert, G.J. Boons, Adaptive immune activation: glycosylation does matter, *Nat. Chem. Biol.* 9 (2013) 776–784.
- [13] C. Zhou, Z. Chen, L. Zhang, D. Yan, T. Mao, K. Tang, T. Qiu, Z. Cao, SEPPA 3.0—enhanced spatial epitope prediction enabling glycoprotein antigens, *Nucleic Acids Res.* 47 (W1) (2019) W388–W394.
- [14] T. Mamedov, J.A. Chichester, R.M. Jones, A. Ghosh, M.V. Coffin, et al., Production of functionally active and immunogenic non-glycosylated protective antigen from *Bacillus anthracis* in *Nicotiana benthamiana* by Co-expression with peptide-N-glycosidase F (PNGase F) of *Flavobacterium meningosepticum*, *PLoS One* 11 (4) (2016) e0153956.
- [15] J. Ghosh, S. Mondal, Nucleic acids and protein content in relation to growth and regression of buffalo (*Bubalus bubalis*) corpora lutea, *Anim. Reprod. Sci.* 93 (2006) 316–327.
- [16] F.A. Soliman, H. Nasr, K. Zaki, Levels of thyroid and thyrotrophic hormones in the blood of Friesian cows at various reproductive stages, *J. Reprod. Fertil.* 6 (1963) 335–340.
- [17] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets, *Mol Biol Evol.* 33 (2016) 1870–1874.
- [18] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent, *J. Biol. Chem.* 193 (1951) 265–275.
- [19] V.L.B.P. Telugu, A.M. Walker, J.A. Green, Characterization of the bovine pregnancy-associated glycoprotein gene family—analysis of gene sequences, regulatory regions within the promoter and expression of selected genes, *BMC Genomics* 10 (2009) 185.
- [20] K. Ushizawa, K. Kaneyama, T. Takahashi, T. Tokunaga, Y. Tsunoda, K. Hashizume, Cloning and expression of a new member of prolactin related protein in bovine placenta: bovine prolactin-related protein-VII, *Biochem. Biophys. Res. Commun.* 326 (2005) (2005) 435–441.
- [21] M. Lotfan, S.A. Ali, M.L. Yadav, S. Choudhary, M.K. Jena, S. Kumar, et al., Genome-wide gene expression analysis of 45 days pregnant fetal cotyledons vis-a-vis non-pregnant caruncles in buffalo (*Bubalus bubalis*), *Gene* 654 (2018) 127–137.
- [22] A.P. Zoli, J.F. Beckers, P. Wouters-Ballman, J. Closset, P. Palmagne, F. Ectors, Purification and characterization of a bovine pregnancy-associated glycoprotein, *Biol. Reprod.* 45 (1991) 1–10.
- [23] A.P. Zoli, J.F. Beckers, F. Ectors, Isolation and partial characterization of a pregnancy associated glycoprotein in sheep (in French), *Ann. Med. Vet.* 139 (1995) 177–184.
- [24] J.M. Garbayo, J.A. Green, M. Manikkam, J.F. Beckers, D.O. Kiesling, A.D. Ealy, R.M. Roberts, Caprine pregnancy-associated glycoproteins (PAGs): their cloning, expression and evolutionary relationship to other PAG, *Mol. Reprod. Dev.* 57 (2000) 311–322.
- [25] H.J. Do, J.H. Kim, L.R. Abeydeera, Y.M. Han, R.L. Matteri, J.A. Green, R.M. Roberts, B.N. Day, R.S. Prather, Expression of pregnancy-associated glycoprotein 1 and 2 genes in *in vivo*, *in vitro* and parthenogenetically derived pre-implantation pig embryos, *Zygote* 9 (2001) 245–250.
- [26] J.A. Green, S. Xie, X. Gan, R.M. Roberts, An aspartic proteinase expressed in the equine placenta, *Adv. Exp. Med. Biol.* 436 (1998) 163–167.
- [27] S. Xie, R.J. Nagel, J.A. Green, J.F. Beckers, R.M. Roberts, Trophoblast-specific processing and phosphorylation of pregnancy-associated glycoprotein-1 in day 15 to 25 sheep placenta, *Biol. Reprod.* 54 (1996) 122–129.
- [28] O.V. Patel, T. Takahashi, K. Imai, K. Hashizume, Generation and purification of recombinant bovine pregnancy associated glycoprotein, *Vet. J.* 168 (2004) 328–335.
- [29] T. Vink, M. Oudshoorn-Dickmann, M. Roza, J.J. Reitsma, R.N. de Jong, A simple, robust and highly efficient transient expression system for producing antibodies, *Methods* 65 (1) (2014) 5–10.

- [30] K. Rousseau, C. Wickstrom, D.B. Whitehouse, I. Carlstedt, D.M. Swallow, New monoclonal antibodies to non-glycosylated domains of the secreted mucins MUC5B and MUC7, *Hybrid Hybridomics* 22 (2003) 293–309.
- [31] B. Forsström, B.B. Axnäs, J. Rockberg, H. Danielsson, A. Bohlin, M. Uhlen, Dissecting antibodies with regards to linear and conformational epitopes, *PLoS One* 10 (2015) e0121673.
- [32] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (2017) W24–W29.
- [33] C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression, *Trends Biotechnol.* 22 (2004) 346–353.
- [34] G. Lithwick, H. Margalit, Hierarchy of sequence-dependent features associated with prokaryotic translation, *Genome Res.* 13 (2003) 2665–2673.
- [35] T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol.* 151 (1981) 389–409.
- [36] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, *Curr. Opin. Biotechnol.* 6 (1995) 494–500.
- [37] D.P. Williams, D. Regier, D. Akiyoshi, F. Genbauffe, J.R. Murphy, Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes, *Nucleic Acids Res.* 16 (1988) 10453–10467.
- [38] A. Li, Z. Kato, H. Ohnishi, K. Hashimoto, E. Matsukuma, K. Omoya, Y. Yamamoto, N. Kondo, Optimized gene synthesis and high expression of human interleukin-18, *Protein Exp. Pur* 32 (2003) 110–118.
- [39] T. Deng, Bacterial expression and purification of biologically active mouse c-Fos proteins by selective codon optimization, *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 409 (1997) 269–272.
- [40] J.A. Kink, M.E. Maley, K.Y. Ling, J.A. Kanabrocki, C. Kung, Efficient expression of the *Paramecium calmodulin* gene in *Escherichia coli* after four TAA-to-CAA changes through a series of polymerase chain reactions, *J. Protozool.* 38 (1991) 441–447.
- [41] O. Barbato, L. Menchetti, N.M. Sousa, A. Malfatti, G. Brecchia, C. Canali, J.F. Beckers, V.L. Barile, Pregnancy-associated glycoproteins (PAGs) concentrations in water buffaloes (*Bubalus bubalis*) during gestation and the postpartum period, *Theriogen* 97 (2017) 73–77.