A 19 kDa protein secreted by the endometrium of the mare is a novel member of the lipocalin family

Ben CROSSETT^{*†‡}, W. R. ALLEN^{*} and Francesca STEWART^{*†‡}

^{*}T.B.A. Equine Fertility Unit, Department of Clinical Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, U.K., and ^{†‡}The Babraham Institute, Babraham, Cambridge CB2 4AT, U.K.

Large quantities of an unusual 19 kDa protein (p19) are secreted into the lumen of the uterus of the mare (*Equus caballus*) during the oestrous cycle and early pregnancy. p19 associates strongly with the acellular capsule that surrounds the young horse conceptus and is believed to be important in maintaining pregnancy. Here we report the complete cDNA sequence encoding p19, its expression patterns in horse tissues and a Southern blot analysis of the gene in horse DNA. The predicted amino acid sequence of the p19 cDNA demonstrated a signal peptide of 18 residues and a mature protein of 162 residues, giving a predicted molecular mass of 18.8 kDa for the secreted protein. Analysis of the predicted amino acid sequence showed that p19 belongs to the lipocalin family of proteins, which are classified on the basis of three conserved amino acid sequence motifs and an eight-stranded anti-parallel β -barrel conformation. Overall, p19 was most similar to the mouse major urinary protein (MUP) lipocalins (30–34% identity). Secondary struc-

tural predictions and structural modelling using the MUP 1 crystal structure coordinates gave convincing evidence that the three-dimensional structure of p19 closely resembles that of the lipocalins. However, p19 has several unusual substitutions in one of the three conserved lipocalin motifs and therefore seems to be a novel member of the family. Lipocalins perform many different functions but most bind small hydrophobic molecules and the majority act as transport proteins. The function of p19 might therefore be as a carrier of a maternal factor needed to sustain the developing embryo during pregnancy, or it could be incorporated into the embryonic capsule and perform some other function. Northern blot analysis demonstrated that expression of p19 is confined to the endometrial lining of the mare's uterus, and hybridization studies *in situ* showed that the mRNA for p19 is localized to the glandular and luminal epithelia of the endometrium. A Southern blot analysis of horse DNA indicated a single gene for p19 that seems to be at least 4.5 kb in size.

INTRODUCTION

Although endometrial proteins are important for embryonic sustenance during early pregnancy in all mammals, two unusual features of early pregnancy in horses make endometrial proteins particularly significant in this species. First, the horse conceptus does not begin to implant and form a stable attachment with the maternal endometrium until as late as 40–42 days after ovulation [1] and secondly, between 6 and 22 days post-ovulation (term is 320–340 days), it is surrounded by an acellular glycoprotein capsule [2–4]. Thus before becoming attached, the conceptus relies on the absorption of endometrial gland secretions (histotrophe) for an exceptionally long period and, for 2–3 weeks, these secretions must pass through the negatively charged capsule. Mares do secrete a protein-rich histotrophe into the uterine lumen [5] and we have partly characterized one of the major components, a small non-glycosylated protein that is secreted in abundance during the late luteal phase (days 10–17) of the oestrous cycle and up to about day 20 of pregnancy [6]. Secretion of the protein seems to be progesterone-dependent, as its presence in uterine flushings correlates with peripheral serum progesterone profiles during the oestrous cycle and its secretion can be induced by exogenous administration of a synthetic progestagen. However, in pregnant mares, the protein becomes undetectable in uterine flushings after about day 20, even though maternal peripheral blood progesterone concentrations remain elevated. This suggests additional control mechanisms for its synthesis in the endometrium. The protein associates strongly with the glycoprotein capsule surrounding the young conceptus and it

therefore probably interacts with, or passes through, the capsule to play a role in maintaining pregnancy. Comparison of the N-terminal 24 amino acids of the protein failed to show significant identity with any other protein sequence in the databases [6].

To investigate this novel protein further, we have cloned the full-length cDNA from an endometrial cDNA library, predicted the probable three-dimensional structure of the protein, examined its expression in horse tissues and performed a Southern blot analysis of horse genomic DNA with the cDNA probe.

MATERIALS AND METHODS

Animals and tissues

Six thoroughbred and four pony mares, aged 3–7 years, were used. Their oestrous cycles were monitored by a combination of teasing with an active stallion for signs of behavioural oestrus and transrectal real-time ultrasound scanning of their ovaries to assess follicular development. To obtain pregnancies, they were mated to a thoroughbred or pony stallion during behavioural oestrus. Ovulation was diagnosed ultrasonographically and confirmed by measuring a rise in progesterone concentrations in peripheral serum samples recovered daily during and beyond oestrus [7]. Day 1 of the cycle or pregnancy was taken as the first day on which progesterone concentrations exceeded 1 ng/ml.

Endometrial biopsies were collected at oestrus (two mares), on day 14 of the cycle and on days 14, 16, 34, 69 and 150 of pregnancy. The biopsies were recovered non-surgically through the cervix in the non-pregnant animals and at days 14, 16 and 34 of pregnancy as described by Ricketts [8]. At days 69 and 150 of

Abbreviations used: MUP, mouse major urinary protein; RATMUP, rat major urinary protein; PP14, placental protein 14; RBP, retinol-binding protein; SCR, structurally conserved regions.

[‡] To whom correspondence should be addressed.

The nucleotide sequence reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number X98459.

Northern blot analysis

Total RNA was extracted as described above and 20 µg of each sample was size-fractionated on a 1% (w/v) agarose gel containing formaldehyde. The gel was washed in a denaturing solution (50 mM NaOH/150 mM NaCl) and then in a neutralizing solution [100 mM Tris (pH 7.0)/150 mM NaCl] before transferring the RNA to a nylon membrane by using positive pressure. The membrane was cross-linked with UV light and pre-hybridized in approx. 5 ml of hybridization buffer (QuickHyb; Stratagene) for 1–2 h at 55 °C. An aliquot (40 ng) of the anti-sense oligonucleotide probe (N7379; see Figure 2) was end-labelled with [³²P]dCTP (ICN, Thame, Oxon., U.K.) by terminal deoxynucleotidyl transferase (Pharmacia Biotech, St. Albans, Herts., U.K.) and added to the membrane in fresh hybridization buffer. Incubation was continued for a further 2 h at 55 °C, after which the membrane was washed four or five times in 1 × SSC [150 mM NaCl/15 mM trisodium citrate (pH 7.0)] with 0.1% SDS at 50 °C and exposed to X-ray film (X-OMAT; IBI-Kodak Ltd., Cambridge, U.K.) for 16 h at –70 °C. The full-length cDNA clone for p19, labelled with ³²P as described under Southern blot analysis, was also used as a probe on one of the Northern blots.

Hybridization *in situ*

Cryostat sections, 7–12 µm thick, were thaw-mounted on microscope slides coated with 3-aminopropyltriethoxysilane (Sigma Chemicals, Poole, Dorset, U.K.), fixed by immersion in a solution of 4% (w/v) paraformaldehyde in PBS (pH 7.2) for 5 min, rinsed in PBS, dehydrated in an ethanol series and stored in ethanol at 4 °C until required. All aqueous solutions were prepared with water that had been stirred overnight with 0.1% (v/v) diethyl pyrocarbonate and then autoclaved. Hybridization was performed in a buffer consisting of 50% (v/v) deionized formamide, 10% (w/v) dextran sulphate, 0.3 M NaCl, 1 × Denhardt's solution [12], 25 mM sodium dihydrogen phosphate (pH 8.0), 20 mM Tris (pH 7.4), 25 mM EDTA (pH 8.0), 50 mM dithiothreitol, 0.5 mg/ml salmon testes DNA and 0.5 mg/ml tRNA. The oligonucleotide probes [N7379 (anti-sense) and N4800 (sense); see Figure 2] were end-labelled with [³⁵S]dATP (Amersham, Bucks., U.K.), diluted in buffer to a concentration of 3 × 10³ d.p.m./µl and 100 µl of this solution was applied to each slide. These were incubated for a minimum of 16 h at 42 °C and then washed in 2 × SSC for 30 min at room temperature and again in 2 × SSC for 45 min at 45 °C. Finally they were rinsed in 2 × SSC for 5 min and 0.1 × SSC for 5 min, dehydrated in an ethanol series, air-dried, coated with a 1:2 dilution of Ilford K5 photoemulsion and exposed for 2 weeks in a light-proof box containing silica gel. The slides were then developed in Ilford LC29 developer, fixed, dried and stained with haematoxylin and eosin.

Southern blot analysis

Lymphocytes were isolated from approx. 200 ml of horse blood by density gradient centrifugation with sodium diatrizoate polysucrose gradients (Accuspin-Histopaque-1077; Sigma) and the DNA was extracted and purified with a Puregene DNA isolation kit (Flowgen, Sittingbourne, Kent, U.K.). Aliquots containing 20 µg of DNA were digested overnight with restriction enzymes (all from Stratagene) and run on a 0.8% agarose gel for 16 h. The gel was then washed once in 0.2 M HCl, once in water, twice in 0.5 M NaOH and once in a neutralizing wash, after which the DNA was capillary-blotted on a nylon membrane. The DNA was cross-linked to the membrane by exposure to UV light and then pre-hybridized for 2 h at 65 °C in Church buffer [200 mM

Na₂HPO₄/50 mM NaH₂PO₄/7% (w/v) SDS]. The p19 576 bp cDNA probe (nt 89–665) was labelled with [³²P]dCTP by random priming [10] and added at a concentration of approx. 200 000 c.p.s./ml in fresh Church buffer. After hybridization for 6–16 h at 65 °C, the membrane was washed several times in 0.5 × SSC/0.1% SDS and then exposed to X-ray film for 1–7 days at –70 °C.

RESULTS

cDNA cloning and sequencing

Although the initial reverse transcriptase-PCR amplification of endometrial mRNA failed (see the Materials and methods section), the oligonucleotide mixture (M1582) based on the N-terminal sequence of p19 gave such a strong signal on Northern blot analysis that it was decided to use M1582 as a probe to clone the full-length cDNA for p19. After amplification, the endometrial cDNA library titre was 1.178 × 10¹⁰ plaque-forming units/ml. Initial screening with M1582 failed to detect any positive clones but by using PCR with the T3 and M1582 primers a 146 bp 5' sequence of p19 was obtained. The sequence consisted of a short 5' untranslated region and an open reading frame corresponding to a potential 18-residue signal peptide, followed by residues 1–20 of the 24-residue N-terminal sequence of p19 (see Figure 1).

The PCR product provided the precise DNA sequence for the 5' region of p19 so that a specific primer could be made to amplify the remainder of the cDNA sequence. A 28-base oligonucleotide (N2782; see Figure 1) was made to a region coding for the first seven residues of the N-terminal sequence. Using the N2782 and T7 primers, a 662 bp PCR product was amplified, cloned, sequenced in both directions and, when combined with the short 5' 146 bp sequence, provided the full coding sequence for p19. However, because the PCR technique can give rise to significant errors during amplification, it was decided that several full-length cDNA sequences should be obtained by screening the cDNA library with the 662 bp PCR product. Over 200 positive plaques were detected in the primary screen of 240 000 plaques, 10 of which were selected for secondary and tertiary screening. Two of these contained full-length cDNA clones that were isolated and sequenced in both directions. Both clones agreed with the coding sequence obtained from the two PCR products; the only difference between the two cDNA clones was that one was slightly longer, both in the 5' untranslated region and in the poly(A) tail region. The nucleotide and translated amino acid sequences of the shorter (737 bp) p19 cDNA clone are shown in Figure 2.

Analysis of the predicted p19 amino acid sequence

Using the pI/molecular mass program on the ExPASy Molecular Biology Server at Geneva University Hospital, the theoretical relative molecular mass for the mature 162-residue protein was 18 793 Da and the theoretical pI was 9.72. The four Cys residues and one potential N-linked glycosylation site at Asn⁸³-Ser⁸⁵ are shown in Figure 2.

The p19 amino acid sequence was compared with the Swissprot database of protein sequences using the Basic Local Alignment Search Tool (BLAST) algorithm in the BLASTP program version 1.4 [13]. The highest match ($P = 1.2 \times 10^{-34}$; 34% identity) was to a lipocalin, rat α_{2u} globulin, also known as rat major urinary protein (RATMUP, Swissprot accession number P02761) and the second highest ($P = 4.9 \times 10^{-34}$; 32% identity) was to mouse major urinary protein V (MUP5; Swissprot accession no. P11591); other members of the MUP group all gave similar

Figure 3 Alignment of the p19 amino acid sequence with members of the lipocalin protein family

Sequence alignment with rat α_{2u} globulin, also known as rat major urinary protein (RATMUP), mouse major urinary proteins I and V (MUP 1 and 5), human placental protein 14 (PP14), pig and horse RBP (pRBP and eRBP) and horse β -lactoglobulin (eBLG). The RATMUP, PP14, pRBP and eRBP lack two to five residues at the N-terminus. The three structurally conserved regions (motifs 1–3) that characterize the protein family are boxed and the most highly conserved residues within motifs 1 and 2 are indicated by arrowheads. The GXW triplet in motif 1 is conserved in p19, whereas the TDY triplet in motif 2, present in 90% of the lipocalins, is not. Two of the Cys residues that have been shown to form a disulphide bond in most of the lipocalins are indicated by linked boxes. See the text for the Swissprot database accession numbers.

scores. Additional members of the lipocalin protein family gave lower scores and included chicken quiescence-specific protein (25% identity; accession no. P21760), rat prostaglandin H2 D isomerase (23% identity; accession no. P22057), mouse neutrophil gelatinase-associated lipocalin (20% identity; accession no. P11672), horse retinol-binding protein (RBP; 20% identity; accession no. U21208), pig RBP (19% identity; accession no. M68860), horse β -lactoglobulin (23% identity; accession no. P07380) and human placental protein 14 (PP14; 19% identity; accession no. P09466). However, unlike the RATMUP and MUP5 identities, these alignments led to several insertions in the p19 sequence. Figure 3 shows the alignment of p19 with several members of the lipocalin family.

The p19 sequence was further analysed by searching the PRINTS database [14], a compendium of protein motif 'fingerprints'. The fingerprint method is a powerful method for detecting conserved structural motifs that characterize protein families and it has been used on several occasions to identify new members of the lipocalin family [15]. Four of the seven structurally conserved regions within the lipocalin family are five residues or fewer in length and have low average conservation indexes, whereas three are 11–18 residues in length and have high conservation indexes [16]. These three structurally conserved regions or motifs are clustered together on one side of the β -barrel, towards the closed end of the barrel. They are used to classify the lipocalin family and are highlighted as boxed areas in the multiple sequence alignment (Figure 3). The most highly conserved residues within the three motifs are the GXW triplet in motif 1 and the TDY triplet in motif 2 (see arrowheads in Figure 3). The p19 sequence scored highly against motif 1 and 3 (including the GXW triplet in motif 1), but not against motif 2. The lack of identity with motif 2 includes the unique substitution of Val-Asp-Pro (VDP) for the Thr-Asp-Tyr (TDY) triplet, which is present in approx. 90% of the lipocalin superfamily [17]. There were also several

Figure 4 Predicted secondary structure of p19

The amino acid sequence of p19 showing the locations of the predicted eight β -strands and two α -helices, which align very well with the known three-dimensional structure of lipocalin. The β -strands form a characteristic barrel shape in all of the solved lipocalin structures, within which a small hydrophobic molecule usually binds. The unique Pro residue in p19 is indicated with an arrowhead.

Figure 5 Northern blot analysis of p19 expression in horse tissues

Northern blot analysis of p19 expression in horse tissues with the oligonucleotide probe (N7379). The probe hybridized with a 1 kb transcript (arrow) in all of the endometrial samples (E) but it was undetectable in the day 14 conceptus (C), liver (L) and all other tissues examined (not shown). The 1 kb transcript was barely detectable in the non-pregnant mare at oestrus (Oest) but was very strong in both of the non-pregnant mares sampled at 14 days after ovulation (one sample shown above). The signal was also very strong at days 14 and 16 of pregnancy (day 16 shown above) but, in keeping with secretion of the protein into the lumen of the uterus, it declined to basal levels later in pregnancy (days 34, 69 and 150 shown above). The three larger transcripts, in the strongest sample, are assumed to be the nascent mRNA and two partly spliced intermediates. The lower panel shows the ethidium bromide-stained RNA before transfer and photographed under UV illumination; the positions of the 28 S and 18 S ribosomal RNA bands are indicated on the blot.

marked differences between p19 and the other lipocalins in this region, which suggests that it is a novel member of the family. The two Cys residues that form an internal disulphide bond in most lipocalins, and therefore almost certainly in p19, are indicated by the linked boxes in Figure 3.

The Protein-Predict program using the Sander and Scheider database of protein family structures (version 1.0) at EMBL [18] was used to predict the secondary structure of p19. It indicated 15% α -helix, 39% β -strand and 46% loop content with eight β -strands and two α -helices (see Figure 4), which aligned very well with the eight β -strands and two α -helices present in the lipocalin

Figure 6 Hybridization analysis *in situ* of p19 expression in horse endometrium

Low-power (c, d) and higher-power (a, b, e, f) bright- and dark-field photomicrographs of frozen sections of endometrium hybridized with the ^{35}S -labelled anti-sense (N7379; a–e) and sense (N4800; f) oligonucleotide probes for p19. (a, b) Images of a section of endometrium from a non-pregnant mare (in oestrus) hybridized with the anti-sense probe. Note the absence of a specific signal in this tissue. (c, d) Images of a section of endometrium from a pregnant mare 16 days after ovulation (anti-sense probe), demonstrating the very strong signal in the glandular endometrium (GE) and weaker signal in the luminal epithelium (LE). (e) A high-power bright-field image of several endometrial glands from another tissue section of the day 16 pregnant mare (anti-sense probe), showing the localization of the signal (black dots) to the epithelial cells lining the glands. (f) A serial section from the same tissue sample hybridized with the sense probe. Scale bar, 10 μm .

structure. In addition it was possible to make a structural alignment of p19, using Swiss-Model version 2.1 at the Glaxo Institute for Molecular Biology in Geneva, Switzerland, with the known crystal structure of MUP 1 [19]. With the model it was possible to confirm that p19 could form the lipocalin fold despite the radical substitution of Pro for Try at position 97 (arrowhead, Figure 4). This substitution gives rise to a sharp turn in the peptide chain, but it does not significantly affect the fold as it occurs in the loop between the sixth and seventh β -strands.

Northern blot analysis

Figure 5 summarizes the Northern blot analysis of the total RNA samples with the anti-sense oligonucleotide probe N7379. The probe hybridized to a 1 kb transcript in all the endometrial samples, consistent with the size of the cDNA clones obtained from the endometrial cDNA library, and, in agreement with our previous observations on secretion of the protein into the uterus [6], levels of the transcript were low at oestrus and at the later stages of pregnancy (days 34, 69 and 150), but high on day 14 of the cycle and days 14 and 16 of pregnancy (see Figure 5). Serum progesterone levels were below 0.5 ng/ml (the limit of sensitivity of the assay) in the two oestrous mares and more than 3 ng/ml in the others. The faint, higher-molecular-mass transcripts in the strongest samples are assumed to represent nascent p19 mRNA and two partly spliced intermediates. RNA from the day 14 conceptus was negative (see Figure 5), as was RNA from the lung, liver, heart, kidney, spleen, skeletal muscle and ovaries recovered from the non-pregnant mare killed on day 14 (Figure 5). The endometrium from this mare gave a very similar signal to that from the other mares on day 14 of the cycle or pregnancy

and her serum progesterone concentration was 3.5 ng/ml. One of the blots was also hybridized to the full-length p19 cDNA clone; it gave an identical result to the oligonucleotide, except that there was more background due to non-specific binding of the probe to the ribosomal RNA bands and to the nylon membrane.

Hybridization *in situ*

When the same anti-sense oligonucleotide probe used in the Northern blots (N7379) was labelled with ^{35}S and hybridized with tissue sections, it showed highly localized hybridization with the glandular and luminal epithelia within the positive endometrial samples, but not with any other tissues (see Figure 6). The signal was very weak in the non-pregnant mares at oestrus (Figures 6a and 6b) and in the pregnant mares on days 34, 69 and 150 after ovulation, but was very strong in the non-pregnant and pregnant mares at days 14 and 16 after ovulation (pregnancy day 16 shown in Figures 6c, 6d and 6e). The control (sense) oligonucleotide probe (N4800) showed no specific binding to any of the tissue sections (Figure 6f).

Southern blot analysis

Figure 7 shows a typical Southern blot obtained after probing the size-fractionated, digested horse DNA with the random-primed, ^{32}P -labelled cDNA p19 probe. Except for the *Pst*I and *Eco*RI digests, which gave two bands, a single band was seen in each lane, suggesting that a single copy of the gene exists in the horse genome. The *Pst*I result was expected because there is a *Pst*I restriction site within the coding region of p19; the *Eco*RI site must therefore be in one of the introns or in a non-translated region not represented in the two cDNA clones. The single bands

Figure 7 Southern blot analysis of horse DNA

Horse lymphocyte DNA (20 μ g per lane) was digested with the restriction enzymes shown and separated on an agarose gel overnight, transferred to a nylon membrane and hybridized with the 32 P-labelled p19 cDNA clone. The positions of the size markers are indicated at the left. *Pst*I was the only restriction enzyme of those used to contain a known site within the p19 coding region and this explains the two bands (arrowheads) seen in this digest. The two bands seen in the *Eco*RI digest were therefore probably due to a restriction site within an intron of the p19 gene.

ranged in size from 4.5 to 10 kb, suggesting that the exons span a region of approx. 4.5 kb. This is also supported by the Northern blot analyses, which indicated a nascent mRNA transcript at approx. 4 kb (see Figure 5).

DISCUSSION

p19 was thought initially to be fetal in origin because it was first detected in the culture medium in which intact horse conceptuses had been incubated overnight at 37 °C. However, additional studies, involving N-terminal sequencing of small amounts of protein isolated from the uteri of mares and from conceptus membranes, indicated that the protein is secreted by the endometrial lining of the uterus of the mare and it is taken up by the glycoprotein capsule that surrounds the horse conceptus [6].

The Northern blot and hybridization *in situ* results presented in this study confirm that p19 is maternal in origin. The mRNA for p19 was undetectable in a day 14 conceptus, but was strongly positive in the mare's endometrium at the expected time in the oestrous cycle and during early pregnancy, in keeping with secretion of the protein into the lumen of the uterus. The Northern blots also confirmed that expression of p19 correlates with progesterone concentrations in the mare's blood during the oestrous cycle but not after day 20 in the pregnant animal, where expression declined to basal levels despite the continuing secretion of progesterone. This suggests that one or more other factors, in addition to progesterone, control the expression of p19 during pregnancy and that the protein is secreted only while the conceptus remains unattached to the endometrium and is surrounded by the acellular capsule. Furthermore, the Northern blot analysis of tissues taken from a day 14 non-pregnant mare, when serum progesterone concentrations were elevated and expression of p19 in the endometrium was also high, showed that p19 mRNA was undetectable in lung, liver, heart, kidney, spleen,

ovary and skeletal muscle. This suggests that the endometrium is the only site of p19 expression in the mare.

Our original estimate of the relative molecular mass of p19, based on SDS/PAGE mobility, was 18 kDa [6], which, given the expected precision of the PAGE method, agrees well with the calculated mass of 18.8 kDa in this study. However, because p19 contains one potential N-linked glycosylation site it could be larger than 18.8 kDa, although a sensitive carbohydrate detection kit was unable to detect any significant carbohydrate content in p19 [6]. Previous SDS/PAGE analyses of mare uterine secretions [5,21] have reported a progesterone-dependent basic protein (U_1) with an estimated molecular mass of 17 kDa, which is almost certainly p19; another recent study [22] assigned a molecular mass of 22 kDa to the same protein.

Although the overall amino acid sequence identities between p19 and the other lipocalins were quite low (19–34% identity), this is not unusual for the family in which the identity between two randomly chosen lipocalin sequences is typically only 20% [15]. This is below the 25% level at which structural similarities frequently go undetected [23]. However, all available crystal structures show the lipocalins to have the same eight-stranded anti-parallel (+1 topology) β -barrel conformation, and most have the three structural motifs highlighted in Figure 3. p19 showed the highest identity (30–34%) with the mouse major urinary proteins (MUPs), which bind pheromones [19]. They are very similar in length: 162 residues for p19, and 162–164 residues for the MUPs. In addition, the two Cys residues that form a disulphide bond in the MUPs are conserved in p19, the shared, identical residues are spread evenly along the molecules, there are no insertions or deletions in the alignments and the predicted eight β -strands and two α -helices in p19 (see Figure 4) show excellent alignment with those in the MUPs.

The lipocalins are a large and growing family of secreted proteins that are between 160 and 190 residues in length, generally bind small hydrophobic molecules within the binding pocket of the cup-shaped β -barrel and often form macromolecular complexes [16]. Most of them are transport proteins [24], although it is now clear that they perform many other important functions, including the control of enzyme synthesis, modulation of cell homeostasis and regulation of immune responses [25]. Although the overall sequence similarity between p19 and the MUPs gave the highest score and p19 is obviously structurally very similar to the MUPs, it is very unlikely that p19 is a member of this group of lipocalins or that it binds a pheromone; p19 is not expressed in the horse kidney and pheromones are not normally found in endometrial secretions. Moreover, the distinctive substitutions in the motif 2 region of p19 (particularly Val-Asp-Pro instead of Thr-Asp-Tyr) also indicate that p19 is not a member of the MUP family and that it is instead a novel lipocalin. These substitutions are situated on a surface loop between the sixth and seventh β -strands in an area that is possibly involved in receptor recognition [16].

The most plausible function for p19 in the mare is as a transport protein carrying a small, maternal product to the conceptus that might be important for embryonic development. An example of such a transport protein is the steroid-dependent lipocalin RBP, a specific carrier for retinol, the principal form in which vitamin A is transported in the serum. Retinoids are endogenous morphogens and are involved in pattern formation during embryogenesis and also in haematopoiesis and differentiation. Consequently, vitamin A concentrations in the developing fetus must be tightly regulated as either insufficient or excessive amounts can result in fetal death, malformation or tumour formation. RBP is important in this regulation and has been identified in the uteri of several species. However, because

p19 has only 20% identity with horse RBP [26], and 23% identity with horse β -lactoglobulin (the retinol transport protein present in milk), p19 cannot be classified as an RBP and it is unlikely to act as an additional carrier of retinol in the pregnant mare.

Another progesterone-dependent lipocalin produced by the female reproductive tract is placental protein 14. This protein was discovered as a member of a large group of soluble human placental proteins [27]. The experimental evidence indicates that PP14, like p19, is synthesized and secreted by the epithelial cells of the endometrial glands [28]. The PP14 gene contains glucocorticoid/progesterone regulatory elements and PP14 secretion is maximal during the late secretory phase of the menstrual cycle [29]. It is also present in the endometrium throughout pregnancy [30] and is believed to have immunosuppressive functions mediated by decreased responsiveness of mitogen- and alloantigen-stimulated lymphocytes to the promoting signals of interleukin 2. However, the sequence identity between PP14 and P18 is only 19%, indicating that they almost certainly perform different functions. Interestingly, although PP14 gives quite high identity scores with the β lactoglobulins (e.g. 44% sequence identity with horse β -lactoglobulin), it has not been found to bind retinol or retinoic acid [31].

The striking correlation between the secretion of p19 into the uterine lumen and the persistence of the embryonic capsule on the horse conceptus raises the possibility that p19 might be incorporated into the capsule and so might account for the measured increase in weight of this unique structure between day 10 and day 18 after ovulation [3]. If so, it could act to aid the transport of other molecules through the capsule or it might perform some other function.

The Southern blot analysis employed in this study indicated that p19 is expressed from a single gene, the exons of which span approx. 4.5 kb. Such a finding is consistent with other members of the lipocalin family, which have a conserved gene structure consisting of seven exons spread over 6–25 kb [32].

We conclude that horse p19 can be classified as a lipocalin on the basis of its structural motifs and that it probably acts as a transport protein during early pregnancy in the mare. Current studies to determine the nature of its possible ligand will be greatly assisted by determining the precise three-dimensional structure of both the recombinant-derived protein produced in bacteria and the native protein purified from uterine flushings. Ultimately we wish to determine the precise function of p19 in the mare and ascertain whether analogues of this interesting protein exist in other mammals.

We thank B. Charleston (Institute of Animal Health, Compton, Berks., U.K.) for stimulating discussions, and W. J. Coadwell (Babraham, Cambridge, U.K.) and

M. J. Stewart (MRC, Cambridge, U.K.) for assistance with computing and for critical appraisal of the manuscript. This work was supported by the Horserace Betting Levy Board (Project 609), The Thoroughbred Breeders' Association of Great Britain and the R. K. Mellon Foundation of America.

REFERENCES

- Samuel, C. A., Allen, W. R. and Steven, D. H. (1974) *J. Reprod. Fert.* **41**, 441–445
- Betteridge, K. J., Eaglesome, M. D., Mitchell, D., Flood, P. F. and Beriault, R. (1982) *J. Anat.* **135**, 191–209
- Oriol, J. G., Sharom, F. J. and Betteridge, K. J. (1993) *J. Reprod. Fert.* **99**, 653–664
- Oriol, J. G., Betteridge, K. J., Clark A. J. and Sharom, F. J. (1993) *Mol. Reprod. Dev.* **34**, 255–265
- Zavy, M. T., Sharp, D. C., Bazer, F. W., Fazleabas, A., Sessions, F. and Roberts, R. M. (1982) *J. Reprod. Fert.* **64**, 199–207
- Stewart, F., Charleston, B., Crossett, B., Barker, P. J. and Allen, W. R. (1995) *J. Reprod. Fert.* **105**, 65–75
- Allen, W. R., Mathias, S., Lennard, S. N. and Greenwood, R. E. S. (1995) *Equine Vet. J.* **27**, 460–464
- Ricketts, S. W. (1975) *Equine Vet. J.* **7**, 102–108
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
- Short, J. M. and Sorge, J. A. (1992) *Methods Enzymol.* **216**, 495–509
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn., Cold Spring Harbour Laboratory, Cold Spring Harbor, NY
- Altshul, S. F., Gish, W., Miller, W., Myers, E. and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Attwood, T. K., Beck, M. E., Bleasby, A. J. and Parry-Smith, D. J. (1994) *Nucleic Acids Res.* **22**, 3590–3596
- Flower, D. R., North, A. C. T. and Attwood, T. K. (1993) *Prot. Sci.* **2**, 753–761
- Flower, D. R. (1995) *J. Mol. Rec.* **8**, 185–195
- Flower, D. R., Sansom, C. E., Beck, M. E. and Attwood, T. K. (1995) *Trends Biochem. Sci.* **20**, 498–499
- Rost, B. and Sanders, C. (1993) *J. Mol. Biol.* **232**, 584–599
- Bocskai, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggoni, A., Findlay, J. B. C. and North, A. C. T. (1992) *Nature (London)* **360**, 186–188
- Reference deleted
- Hinrichs, K., Kenney, R. M. and Sharp, D. C. (1989) *Equine Vet. J. Suppl.* **8**, 49–55
- Beier-Hellwig, K., Kremer, H., Bonn, B., Linder, D., Bader, H. and Beier, H. M. (1995) *Reprod. Dom. Anim.* **30**, 295–298
- Sander, C. and Schneider, R. (1991) *Proteins Struct. Funct. Genet.* **9**, 56–68
- Pervais, S. and Brew, K. (1985) *Science* **228**, 335–337
- Flower, D. R. (1994) *FEBS Lett.* **354**, 7–11
- McDowell, K. J., Adams, M. H., Franklin, K. M. and Baker, C. B. (1995) *Biol. Reprod.* **52**, 438–443
- Bohn, H., Kraus, W. and Winckler, W. (1982) *Placenta (suppl.)* **4**, 67–81
- Julkunen, M., Koistinen, R., Suikkari, A. M., Seppälä, M. and Jänne, O. A. (1990) *Mol. Endocrinol.* **4**, 700–707
- Julkunen, M., Koistinen, R., Sjöberg, J., Rutanen, E., Wahlstrom, T. and Seppälä, M. (1986) *Endocrinology (Baltimore)* **118**, 1782–1786
- Seppälä, M., Riittinen, L., Kämäräinen, M. and Julkunen, M. (1992) *Reprod. Fert. Dev.* **4**, 611–618
- Seppälä, M., Riittinen, L., Julkunen, M., Vartiainen, E. and Huhtala, M. L. (1988) *J. Reprod. Fert. Suppl.* **36**, 127–141
- Igarashi, M., Nagata, A., Toh, H., Urade, Y. and Hayaishi, O. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5376–5380