Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 2456–2460, March 1991
Biochemistry

Multiple forms of mRNA encoding human pregnancy-associated endometrial α_2-globulin, a β-lactoglobulin homologue

(endometrial proteins/cDNA sequences/alternative splicing)

J. GARDE*†‡, S. C. BELL*†, AND I. C. EPERON*

*Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom; and †Department of Obstetrics and Gynaecology, University of Leicester, Clinical Sciences Building, Leicester Royal Infirmary, Leicester LE2 7LX, United Kingdom

Communicated by Neal L. First, September 25, 1990 (received for review May 18, 1990)

ABSTRACT **Human pregnancy-associated endometrial α_2-globulin (α2-PEG) is the major secretory protein product of the endometrium during embryo implantation and the first few weeks of pregnancy. It is a homologue of β-lactoglobulin, a retinol binding protein, but unlike β-lactoglobulin it is not found in the mammary gland. The cloning and sequencing of 34 α2-PEG clones has revealed several minor variant forms indicative of alternatively spliced α2-PEG pre-mRNA. These minor forms have also been detected amongst uncloned cDNA after PCR amplification. Some of these mRNAs would give rise to forms of α2-PEG protein lacking internal sequences, whereas others affect the mRNA sequences on the 3′ boundary of the presumed termination codon. Sequences within the cDNA clones are consistent with the existence of splice sites, and together with similarities found between α2-PEG cDNA and β-lactoglobulin gene sequences there is good evidence in support of an unusual scheme for the alternative splicing of α2-PEG pre-mRNA involving both alternative 5′ splice sites and alternative 3′ splice sites. This scheme suggests that the α2-PEG and β-lactoglobulin genes share a similar structure in at least two regions, and it is likely that β-lactoglobulin pre-mRNA would show a similar pattern of alternative splicing for one of these regions.**

The human endometrium secretes a number of proteins whose levels are modulated during the menstrual cycle and pregnancy. The major product seen *in vitro* during the mid to late luteal stage of the menstrual cycle and the first trimester of pregnancy has been named human pregnancy-associated endometrial α_2-globulin (α2-PEG) (1, 2). It has been localized immunochemically to the secretory glandular epithelium of this tissue, where it is thought to be produced in response to progesterone-dependent differentiation of the endometrium (3, 4). α2-PEG has also been detected at high concentrations in amniotic fluid (5) and in uterine luminal fluid (6), where its temporal profile reflects changes in the *in vitro* rate of synthesis and secretion of α2-PEG by the endometrium (1, 2). This suggests that movement into these compartments is the major secretory route for α2-PEG. Serum concentrations of α2-PEG are low (≈2% of the amniotic fluid concentrations) but show a similar pattern of change, indicating their potential clinical use as a marker of endometrial function (7, 8).

α2-PEG is a 56-kDa homodimeric glycoprotein whose subunits exhibit microheterogeneity of molecular mass and pI (9, 10). The main clue as to its function has been the demonstration of significant amino acid sequence similarity to β-lactoglobulin, a protein expressed in the glandular epithelium of the mammary gland in ruminants and certain other species (11). β-Lactoglobulin is itself homologous to human retinol binding protein and has been shown to bind retinol as well as other small hydrophobic molecules (12, 13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Several other proteins have been identified that are very similar to α2-PEG, which include placental protein 14 (PP14), chorionic α_2-microglobulin, progestagen-dependent endometrial protein, and α-uterine protein (10). However, differences in N-terminal sequence analysis (14–16), apparent molecular mass (10), and immunocytochemical localization have been reported (10). Such findings suggest that polymorphic forms of α2-PEG may exist.

If polymorphic forms of α2-PEG do exist, their identification may assist in understanding the function of this protein. Therefore, we have investigated whether polymorphism occurs at the primary structure level by isolating multiple α2-PEG cDNA clones and performing detailed sequence analysis on each.§

MATERIALS AND METHODS

Preparation of Endometrial cDNA Libraries. Individual first trimester endometrial tissue samples were incubated overnight at 37°C in culture medium containing [^{35}S]methionine, and the α2-PEG protein content of the supernatants was assessed by NaDodSO$_4$/PAGE (17). RNA was extracted from those endometrial tissues that secreted the highest levels of α2-PEG (18), and translation *in vitro* was used to identify samples with the highest levels of α2-PEG mRNA. cDNA was prepared (19) from two such samples and cloned into Lambda ZAP II (Stratagene).

Isolation and Sequencing of α2-PEG cDNA Clones. The two Lambda ZAP II libraries were converted to plasmid libraries in pBluescript by the automatic excision from λ sequences as described by Stratagene. A 5′ 231-base-pair (bp) *Pst* I fragment from an ovine β-lactoglobulin cDNA (20) was labeled by random priming (21) and used to screen the colonies by standard hybridization procedures. In addition, a 39-mer oligonucleotide, identical in sequence to the 5′ end of the mature protein coding region of PP14 cDNA, was labeled by primed synthesis and used as an alternative probe for screening the colonies. Single-stranded DNA was prepared from pure colonies, and the nucleotide sequence of the cDNA from each clone was analyzed by the dideoxynucleotide chain-termination method (22).

Analysis of Uncloned α2-PEG cDNA. The uncloned cDNA was amplified by the PCR. A pair of 20-mer or 33-mer oligonucleotides derived from the α2-PEG cDNA sequences was used (see text). Each reaction mixture (50 μl) contained cDNA (10 ng) in 250 mM Tris·HCl, pH 7.5/250 mM $(NH_4)_2SO_4$/bovine serum albumin (DNase free) (2 mg/ml)/2-mercaptoethanol (0.0007%)/2 mM or 3.5 mM $MgCl_2$/250

Abbreviations: α2-PEG, pregnancy-associated endometrial α_2-globulin; PP14, placental protein 14; nt, nucleotide(s).

‡Present address: Department of Genetics, University of Liverpool, P.O. Box 147, Liverpool L693BX, United Kingdom.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M61886).

μM each deoxynucleotide triphosphate/10 pmol of each primer/4 units of *Thermus aquaticus* DNA polymerase (*Taq* polymerase) (Amersham). The reactions were performed for 30 cycles, each cycle including denaturation at 92°C for 1.5 min, annealing at 40°C for 1.5 min, and elongation at 72°C for 1.5 or 2 min. The PCR products were fractionated on low melting point agarose gels containing ethidium bromide and were then subjected to analysis by restriction enzyme digestion or Southern blots.

Transcription, Translation, and Immunoprecipitation. The DNA from pBluescript/α2-PEG cDNA clones (1 μg) was linearized with *Tth* III/I or *Bal* I, and capped sense transcripts were synthesized in the presence of 0.5 mM m^{7}GpppG by using T7 or T3 RNA polymerase (depending on the orientation of the cDNA) (23). The transcripts (0.1 μg) were translated in rabbit reticulocyte lysate with [^{35}S]methionine as the labeled amino acid (Amersham), and some of the translated peptides were immunoprecipitated with an IgG fraction derived from an α2-PEG antiserum and nonimmune serum (24). Endometrial proteins were labeled by the growth of the tissue in medium containing [^{35}S]methionine (17).

RESULTS

Isolation of cDNA Clones. Two human cDNA libraries were prepared in Lambda ZAP II from the endometrial poly(A) RNA of two individuals preselected for high α2-PEG mRNA content. The basis of the selection was the detection of abundant α2-PEG polypeptide after *in vitro* translation of this RNA and the detection of abundant α2-PEG protein in culture medium after incubation with a sample of the endometrial tissue from which this RNA was extracted (data not shown). The Lambda ZAP II libraries were converted to pBluescript plasmid libraries, which reduced both the vector size [from 48 to 2.96 kilobases (kb)] and the chance of cross-hybridization of DNA probes to the vector. Two probes were used. An oligonucleotide was prepared that contained sequences from the N-terminal region of α2-PEG, based on the reported amino acid sequences of α2-PEG and PP14 (14, 16, 25) and the cDNA sequence of PP14 (15). In addition, we were concerned that variants might be missed by this approach, and so we used a 231-nucleotide (nt) fragment of an ovine β-lactoglobulin cDNA; in the region of the available N-terminal sequence of 38 amino acids of α2-PEG (14), β-lactoglobulin has 17 identical amino acids, and over the 231 nt the fragment is 78% identical with the aligned region of PP14 cDNA (15). In total, 34 positive clones were isolated by using one or the other of the probes.

Nucleotide Sequence of α2-PEG cDNA and the Deduced Amino Acid Sequence. The nucleotide sequence was determined on one strand of each of two cDNA clones in opposite orientations. The clones chosen contained the longest 5′ and 3′ ends. The α2-PEG nucleotide sequence and the deduced amino acid sequence for the protein are shown in Fig. 1. The two strands sequenced were completely complementary, one being just 2 bp longer than the other at the 5′ end of the cDNA. The 49 nt at positions 579–627 (Fig. 1) were found in a rare cDNA and did not occur in the two standard clones from which the remaining sequence was obtained.

The derived sequence for standard α2-PEG cDNA is identical to PP14 cDNA apart from two minor differences. One of these differences is the change from a glycine codon, GGG (PP14), to a glutamic acid codon, GAG (α2-PEG), at residue 77, which has also been noted in three partial PP14 cDNA clones (15). The other difference is the change from TT (PP14) to CG (α2-PEG) at position 781/782 in the 3′ noncoding region.

α2-PEG cDNA Clones Include Minor Variant Types. Single track DNA sequences were compared over the entire length of one strand of each of the 34 cDNA clones isolated. This indicated that only 19 were of the standard type (Fig. 1 minus nt 579–627). More detailed sequence analysis revealed that the remainder comprised a number of variant types defined by blocks of missing or additional DNA compared to the standard cDNA. The coding region deletions preserved the reading frame. The structure of these variants and the number of clones found of each type are shown in Fig. 2.

Uncloned α2-PEG cDNA Contains Minor Variant Types. The possibility that the variant types of α2-PEG cDNA arose from cloning artifacts can be ruled out if they can be shown to occur among uncloned cDNA. For this purpose, α2-PEG cDNA present among the uncloned cDNA used to make the libraries was subjected to amplification by the PCR with primers that anneal to the 5′ and 3′ noncoding sequences of standard α2-PEG cDNA (nt 16–35 and 628–647; Fig. 1). The dominant amplification products seen after agarose gel electrophoresis included two doublets corresponding to lengths of ≈580 and ≈300 bp and a dominant band of ≈100 bp. When the doublet DNA bands (≈580 and ≈300 nt) were eluted, 5′-end-labeled, and fractionated on an acrylamide gel, the upper band in each doublet was seen to be 29 nt longer than the lower band, and the four amplified products were pre-

```
                                         -18                                        -10                                      1
                                         Met Leu Cys Leu Leu Leu Thr Leu Gly Val Ala Leu Val Cys Gly Val Pro Ala Met Asp Ile Pro
  1     GGCTCCAGAGCTCAGAGCCACCCACAGCCGCAGCC ATG CTG TGC CTC CTG CTC ACC CTG GGC GTG GCC CTG GTC TGT GGT GTC CCG GCC ATG GAC ATC CCC
                      10                  ⇓                20                                        30                          ⇓
    Gln Thr Lys Gln Asp Leu Glu Leu Pro Lys Leu Ala Gly Thr Trp His Ser Met Ala Met Ala Thr Asn Asn Ile Ser Leu Met Ala Thr Leu Lys
102 CAG ACC AAG CAG GAC CTG GAG CTC CCA AAG TTG GCA GGG ACC TGG CAC TCC ATG GCC ATG GCG ACC AAC AAC ATC TCC CTC ATG GCG ACA CTG AAG
          40                                  50                                  60
    Ala Pro Leu Arg Val His Ile Thr Ser Leu Leu Pro Thr Pro Glu Asp Asn Leu Glu Ile Val Leu His Arg Trp Glu Asn Asn Ser [Cys] Val Glu
198 GCC CCT CTG AGG GTC CAC ATC ACC TCA CTG TTG CCC ACC CCC GAG GAC AAC CTG GAG ATC GTT CTG CAC AGA TGG GAG AAC AAC AGC [TGT] GTT GAG
     70                                  80                                  90                                        100
    Lys Lys Val Leu Gly Glu Lys Thr Glu Asn Pro Lys Lys Phe Lys Ile Asn Tyr Thr Val Ala Asn Glu Ala Thr Leu Leu Asp Thr Asp Tyr Asp
294 AAG AAG GTC CTT GGA GAG AAG ACT GAG AAT CCA AAG AAG TTC AAG ATC AAC TAT ACG GTG GCG AAC GAG GCC ACG CTG CTC GAT ACT GAC TAC GAC
                                   ⇓   110                                 120                                 130
    Asn Phe Leu Phe Leu [Cys] Leu Gln Asp Thr Thr Thr Pro Ile Gln Ser Met Met [Cys] Gln Tyr Leu Ala Arg Val Leu Val Glu Asp Asp Glu Ile
390 AAT TTC CTG TTT CTC [TGC] CTA CAG GAC ACC ACC ACC CCC ATC CAG AGC ATG ATG [TGC] CAG TAC CTG GCC AGA GTC CTG GTG GAG GAC GAT GAG ATC
                               140                                 150                                 160
    Met Gln Gly Phe Ile Arg Ala Phe Arg Pro Leu Pro Arg His Leu Trp Tyr Leu Leu Asp Leu Lys Gln Met Glu Glu Pro [Cys] Arg Phe ***⇓
486 ATG CAG GGA TTC ATC AGG GCT TTC AGG CCC CTG CCC AGG CAC CTA TGG TAC TTG CTG GAC TTG AAA CAG ATG GAA GAG CCG [TGC] CGT TTC TAG GTG
                                 ⇓                     ⇓
582 AGCTCCTGCCTGGTC CTGCCTCCTGGGTGA CCTGTAAACCCAACA GCTCACCTCCGCCTC CAGGAAGACCAGACT CCCACCCTTCCACAC CTCCAGAGCAGTGGG ACTTCCTCCTGCCCT

702 TTCAAAGAATAACCA CAGCTCAGAAGACGA TGACGTGGTCATCTG TGTCGCCATCCCCTT CCTGCTGCACACCTG CACCACGGCCATGGG GAGGCTGCTCCCTGG GGGCAGAGTCTCTGG

822 CAGAGGTTATTAATA AACCCTTGGAGCATG
              oooo oo
```

FIG. 1. Nucleotide sequence of the longest form of α2-PEG. This sequence was obtained from sequencing the opposite strands of two separate clones. The 5′ terminus of one clone corresponds to nucleotide number 1, and the 3′ terminus of the other corresponds to nucleotide number 3. The two strands were otherwise entirely complementary. The deduced α2-PEG amino acid sequence is shown above the DNA, with the negative numbers referring to the signal peptide and the positive numbers referring to the mature protein. The underlined regions correspond to the protein sequences we have reported (12). The four cysteine residues are boxed, and the consensus polyadenylylation signal is indicated by a series of open circles. The sequence shown includes a 49-bp insert (positions 579–627), which was present in another α2-PEG clone and not in the two described above. Arrows refer to the positions of sequences missing in one or more further clones; positions 579–627 (49 bp) (missing in most α2-PEG clones), positions 607–627 (21 bp), positions 132–197 (66 bp), positions 132–413 (282 bp) (see text).

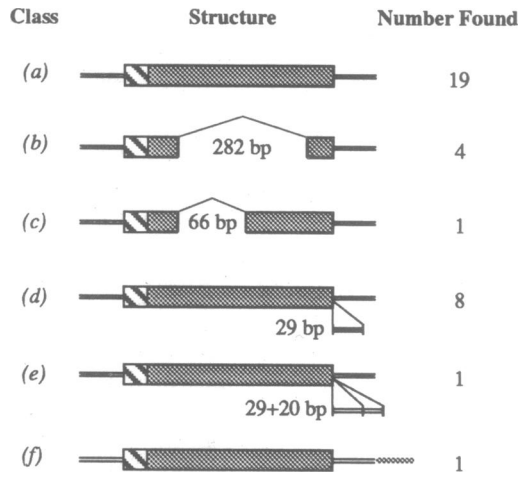

FIG. 2. The structure and frequency of different classes of α2-PEG cDNA clones. Single-track sequencing on one DNA strand was carried out over the entire length of each cDNA clone, and four-track sequencing was carried out over all the regions where the sequence differed from that of standard cDNA (class a), either on one strand or on both strands where variant clones in opposite orientation were available (class d). Two clones lacked a poly(A) tail, many also lacked between 1 and 35 nt from the 3′ end, and many lacked between 1 and 32 nt from the 5′ end compared to the cDNA sequence shown in Fig. 1. The 29 bp of class e are the same as those in class d. The 130 bp of class f are probably from pre-mRNA that had undergone splicing but not 3′ end maturation. □, Mature protein coding sequences; □, signal peptide coding sequences; ——, 5′ and 3′ noncoding sequences.

cisely the expected size for standard cDNA, cDNA that contained an additional 29-nt sequence after the stop codon, cDNA that lacked codons 15–108, and cDNA that lacked these codons but also contained the additional 29-nt sequence; this last is a combination that was not found among the cDNA clones analyzed (Fig. 3).

Laser densitometry was used to determine the relative levels of the two bands in each lane of Fig. 3 (samples 1 and 2 originate separately from the cDNA used to make the two libraries). The combined ratio of the standard cDNA to the longer form for samples 1 and 2, when it is weighted according to the different numbers of clones analyzed from each cDNA library, is 74:26. The frequencies with which the cDNA clones were found (20:8; Fig. 2) are not significantly different (by χ^2 analysis) from this ratio. The equivalent ratio for the cDNA lacking codons 15–108 is 80:20. The corresponding cDNA clones were found in the ratio 4:0, which, given the small numbers involved, is not inconsistent with the supposition that for bands of similar sizes the PCR amplification reflects the abundance of the original cDNA faithfully.

The identities of the PCR products were confirmed by digestion of the labeled molecules with *Bal* I, *Ava* II, *Mae* I, or *Hae* III. The two pairs of doublets produced the predicted fragments, whereas the 100-bp-long PCR product did not appear to be related to α2-PEG (data not shown). Under different conditions of the PCR, other minor products were formed; these, too, appeared to be forms of α2-PEG cDNA because, unlike the 100-bp band, they gave positive results on blotting (data not shown).

Variant α2-PEG Proteins Are Predicted. If the large and small splice mRNAs are translated *in vivo* they would specify minor low molecular weight α2-PEG protein forms (lacking amino acids 15–108 and 15–36, respectively; Fig. 1). However, such proteins had not been detected during the purification of α2-PEG from the cytosol of human pregnancy endometrium (10). The purification procedure used relied on the detection of α2-PEG in various chromatographic fractions by using an adsorbed α2-PEG antiserum. We therefore

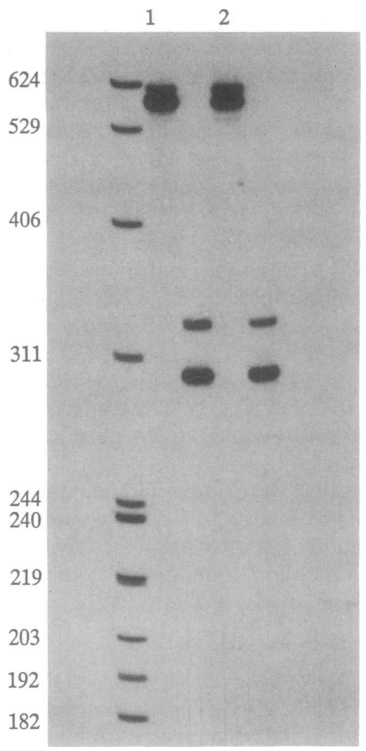

FIG. 3. Analysis of uncloned α2-PEG cDNA after specific amplification by PCR. PCR was performed on uncloned cDNA from two separate individuals (lanes 1 and 2) in buffer containing 2 mM $MgCl_2$; 20-mer primers with annealing sites in the 5′ and 3′ noncoding regions of α2-PEG were used (positions 15–35 and 579–599; see Fig. 1), and the DNA from 10 reactions was pooled and fractionated on a 2% low melting point agarose gel. The ≈580- and ≈310-bp doublet PCR products were eluted from the gel, 5′-end-labeled with [γ-^{32}P]ATP, and purified on an acrylamide gel. The recovered DNA from each doublet band was pooled, and a small portion (≈20 cps) fractionated on a 6% acrylamide gel. Markers are shown on the left with lengths in bp.

tested whether the predicted minor low molecular weight α2-PEG protein forms would be reactive with this antiserum to ascertain whether a lack of reactivity is a possible explanation for the failure to detect them.

Polypeptides were translated from transcripts derived from cDNA forms a–c (Fig. 2). The signals produced on immunoprecipitation with an α2-PEG antiserum IgG fraction were 8- to 10-fold higher than the control values with nonimmune serum (data not shown). Neither IgG fraction precipitated ovalbumin.

Repeated analysis of the proteins secreted by cultured endometrial tissue showed that several small polypeptides of very low abundance could be immunoprecipitated (data not shown). One of these was ≈9.5 kDa, the right size for the small splice mRNA translation product, but it was of disproportionately low intensity relative to α2-PEG (compared with the ratio of cDNA clones isolated), and it has not been possible to demonstrate that this polypeptide is related to α2-PEG.

DISCUSSION

The work described in this paper was undertaken to investigate the possibility that α2-PEG protein might be polymorphic, comprising a class of closely related polypeptides. The sequences of the most abundant form of cDNA show that α2-PEG is very similar to PP14; the single amino acid difference could be due to either errors in the PP14 sequence (15) or allelic differences. The same explanations can be

invoked for the differences between our cDNA sequences and the N-terminal sequence of the protein reported earlier (14, 16). Likewise, the absence of major α2-PEG mRNA forms that could encode the related proteins chorionic α_2-microglobulin, progestagen-dependent endometrial protein, and α-uterine protein (see Introduction; ref. 10) suggests that the reported differences between these proteins and α2-PEG are methodologic in origin. The microheterogeneity of α2-PEG is probably a consequence of differences in glycosylation. However, a number of variant forms of mRNA were characterized and the possibility remains from the PCR experiments that there are yet more forms.

The forms that have been characterized show identical nucleotide sequences with the exception of apparent deletions or insertions at common sites. Such a pattern is commonly associated with alternative splicing. In the absence of extensive information about genomic sequences it is not usually possible to draw secure inferences about the patterns of alternative splicing. However, there are two important pieces of evidence from our analysis, which show that alternative splicing is certainly the explanation for the cDNA sequences found and also suggest a viable scheme for splicing.

The first important point is that in every case one end of the deletion or insertion maps within the longest cDNA at a sequence that fits very closely to the consensus for a 5′ or 3′ splice site (26–28). This is shown in Fig. 4*A*. Calculations of the significance of this match are comparatively easy for 5′ splice site sequences, where the consensus sequence covers a clearly defined region of 9 nt [(C/A)AGGU(A/G)AGU]. If it is assumed that in a random RNA sequence all 4 nt are present in equal proportions, then the probability that a sequence would be found exactly at one end of the deletion with a match to the consensus as good as or better than that at position 578/579 is ≈0.1%: the corresponding probability for the site at position 607/608 is ≈1.2% (binomial calculations with the constraint that the GT is absolutely required). Clearly, the probability that both positions would by chance match the consensus to that degree is minimal.

The second piece of evidence is derived from alignment of the α2-PEG cDNA sequence with that for ovine β-lactoglobulin. The nucleotide protein coding sequences of the two cDNAs are 67% identical, with perfect alignment; the 3′ untranslated portions are identical in 33% of nucleotides and they differ in length by only 21 nt (15). The arrangement of ovine β-lactoglobulin exons is known, together with the sequences across the splice sites (29, 30). Strikingly, the common 5′ end of the coding region deletions falls exactly at a splice site in the β-lactoglobulin gene. Although the 3′ junction of the 3′ noncoding region variant sequences cannot be aligned with certainty with the corresponding exon of β-lactoglobulin, one of the three alternative 5′ splice sites is used as such in β-lactoglobulin (Fig. 4*B*). This suggests the scheme for the alternative splicing of α2-PEG shown in Fig. 5.

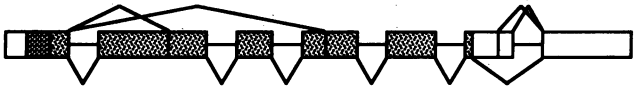

FIG. 5. Predicted splicing patterns for the α2-PEG gene. Arrangement of the α2-PEG gene is based on that of the aligned ovine β-lactoglobulin gene (24, 30). Exons are boxed. □, 5′ and 3′ noncoding sequences; ▨, signal peptide coding sequences; ▩, mature protein coding sequences.

A question arises about the likelihood of the sequences identified above acting as splice sites—i.e., whether the existence of the sequences supports our contention that they are splice sites. In practice, there are often a number of apparently viable sequences in genes that are not used as such (27, 31). However, for 5′ splice sites it has been shown in previous work from this laboratory that there is a significant correlation between the extent to which a splice site sequence resembles the consensus sequence (measured in a variety of ways) and its intrinsic strength as assayed by preferences *in vivo* between potential alternative sites (32, 33). The candidate 5′ splice sites in the case of α2-PEG are those created by insertions of 29 or 49 nt next to the termination codon (Figs. 2 and 4).

Although the correlation described above is not sufficient to allow the strength of any candidate splice site sequence not ranked in our hierarchy to be predicted with certainty (33), it can be said that the site at position 578/579 is likely to be a strong site and that at 607/608 is likely to be relatively weak. Obviously, the intrinsic strength of a site is modified considerably by surrounding sequences and the relative positions of alternative sites (34–37), but it is nonetheless striking that use of the putative site that would be expected to be stronger is clearly preferred (25:8 among the cDNA clones). The presence of extra introns between the three alternative 5′ splice sites is made less likely by the alignment with β-lactoglobulin. It is interesting that the 607/608 site aligns with the 5′ splice site identified in β-lactoglobulin (Fig. 4), and that in β-lactoglobulin the sequence reveals a corresponding potential site equivalent to that at 578/579 in α2-PEG: in β-lactoglobulin, the observed 607/608 site and the predicted site at 578/579 fits the consensus sequence to an equivalent degree.

The situation is less clear for the candidate 3′ splice sites involved in the coding region deletions (codons 15–36 and 15–108). Both of these deletions preserve the reading frame and end in the dinucleotide sequence AG (found at the 3′ splice site of all eukaryotic pre-mRNA introns). Two other sequence elements near the 3′ splice site appear to be involved in splicing: a branch site and a polypyrimidine tract. Some systematic studies have been reported for preferences

A

5′ splice site			branch site			3′ splice site		
578/579	UAG/GUGAGC	88.2	167-173	CAACAA*C		197/198	UGGCGACACUGAAG/G	
607/608	UGG/GUGACC	68.9	377-383	UACUGA*C		413/414	UUCUCUGCCUACAG/G	
consensus	CAG/GUAAGU A G	100	consensus	YNCURA*Y		consensus	YYYYYYYYYNYCAG/G	

B

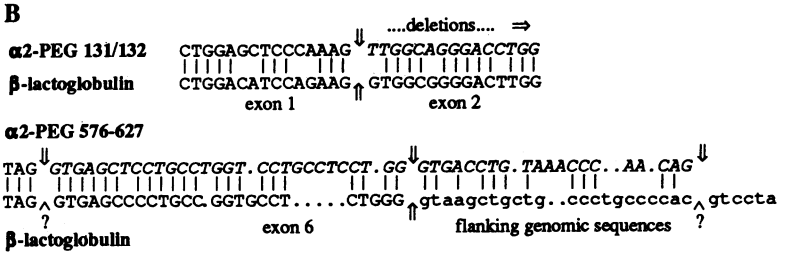

FIG. 4. Alignments between deletion/insertion boundaries and splice site consensus sequences or splice sites in ovine β-lactoglobulin. (*A*) Putative splice site sequences in α2-PEG cDNA. The numbers shown to the right of the 5′ splice sites are the averages, relative to the consensus, of the values for the frequency (percent) with which these nucleotides are found at these positions in primate splice sites (27). The branch site nucleotide is followed by an asterisk. (*B*) Alignment of the boundaries for α2-PEG insertions or deletions with exon junctions (upper alignment) or splice sites (lower alignment) in ovine β-lactoglobulin cDNA and genomic sequences, respectively. Question marks point to positions in β-lactoglobulin that are homologous to the other probable 5′ splice sites in α2-PEG.

in vivo among sequences in the region of the branch site consensus (38–40). The best prospective branch site sequence for the splice deleting codons 15–108 is UACUGAC, which is known to function relatively well (39). There are no good candidates for the corresponding site required for the splice deleting codons 15–36, the best being CAACCAAC, which has not been tested in mammalian systems and is unlikely to be a favored sequence. Given that branch site sequences can affect the use of alternative splice sites (38, 39), it may be significant that the two deletions were found in four and one cDNA clones, respectively. There have been no experiments reported on the competitive strength *in vivo* of a range of 3′ splice sites (polypyrimidine tract and AG), although some aspects of the dependence on primary sequence can be inferred from studies *in vitro* (41, 42). The two polypyrimidine tracts for the candidate sites are short but within the range found *in vivo*; that for the splice deleting codons 15–36 appears to be the shorter, but it appears to be in a more favorable position, close to the branch site and lacking intervening blocks of purines (41, 42). Thus, no predictions can be made of the relative efficiency of these elements, but all the available evidence is consistent with the positions and relative levels of use of the splice sites proposed in Fig. 5. The difficulty of demonstrating the existence of the polypeptides expected from these forms of mRNA suggests that translation is poor or that the proteins are unstable.

A scheme such as that proposed in Fig. 5 is most unusual. Cellular (as opposed to viral) genes with alternative 5′ splice sites or 3′ splice sites within exons are rare, and it is even more unusual to find evidence for both patterns within one gene. We are aware of only one other example: the gene for the 70-kDa protein of U1 small nuclear ribonucleoproteins (43).

The alignment with ovine β-lactoglobulin in the 3′ untranslated portion of the gene has some interesting implications. We have pointed out already that the preferred 5′ splice site in α2-PEG is adjacent to the termination codon and that the same sequence is present in ovine β-lactoglobulin. Thus, we would expect to find some evidence for alternative splicing here. The sequence reported for a bovine β-lactoglobulin cDNA appears to support this; it is very similar to the ovine sequence (44), with the important exception that it lacks the 25-nt sequence homologous to α2-PEG beyond the termination codon. The full sequence of the most downstream 5′ splice site in α2-PEG is unknown, but it seems unlikely from inspection of the β-lactoglobulin genomic sequence that the corresponding GT would act as a 5′ splice site even at the level found in α2-PEG (1 in 34 cDNA clones) (Fig. 4).

The sequences 3′ to the termination codon are highly conserved between human α2-PEG and ovine β-lactoglobulin; indeed within the 29-nt block they are more conserved than the coding sequences. The slight deletions or insertions required to align the two cDNAs suggest that this region is unlikely to be used for translation (after, for example, a frameshift prior to the termination codon), and it is most likely that this sequence plays a role in determining the stability or, possibly, location of the mRNA. The 29-nt block is rich in G and C (20/29 nt), in contrast to the known instability determinant of a number of mRNAs (45, 46), and thus one possibility is that it may be expressed differentially at various stages in endometrial development to elevate the level of mRNA and protein. Further studies will be directed toward clarification of the significance of these different forms of mRNA and protein in terms of the molecular mechanism of endocrine regulation of α2-PEG expression and, ultimately, the biological function of α2-PEG.

We thank S. Ali and A. J. Clark for providing the β-lactoglobulin clone used in this study and Jill Ashmore for her technical assistance. This work was supported by the Wellcome Research Trust.

1. Bell, S. C., Patel, S. R., Kirwan, P. H. & Drife, J. O. (1986) *J. Reprod. Fertil.* **77**, 221–231.
2. Bell, S. C. (1986) *Hum. Reprod.* **1**, 129–143.
3. Waites, G. T., Wood, P. L., Walker, R. A. & Bell, S. C. (1988) *J. Reprod. Fertil.* **82**, 665–672.
4. Waites, G. T. & Bell, S. C. (1989) *J. Reprod. Fertil.* **87**, 291–300.
5. Bell, S. C., Hales, M. W., Patel, S. R., Kirwan, P. H., Drife, J. O. & Milford-Ward, A. (1986) *Br. J. Obstet. Gynaecol.* **93**, 909–915.
6. Bell, S. C. & Dore-Green, F. (1987) *J. Reprod. Immunol.* **11**, 13–29.
7. Bell, S. C., Jackson, J., Dore-Green, F., Kirwan, P. H. & Drife, J. O. (1987) *Hum. Reprod.* **2**, 384–387.
8. Wood, P. L., Walker, R. A. & Bell, S. C. (1989) *Hum. Reprod.* **4**, 140–146.
9. Bell, S. C., Hales, M. W., Patel, S., Kirwan, P. H. & Drife, J. O. (1985) *Br. J. Obstet. Gynaecol.* **92**, 793–802.
10. Bell, S. C. (1986) *Hum. Reprod.* **1**, 313–318.
11. Godovac-Zimmermann, J. & Braunitzer, G. (1987) *Milchwissenschaft* **42**, 294–296.
12. Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadorao, R., Jones, T. A., Newcomer, M. E. & Kraulis, P. J. (1986) *Nature (London)* **324**, 383–385.
13. Fugate, J. E. & Song, P. S. (1980) *Biochim. Biophys. Acta* **625**, 28–42.
14. Bell, S. C., Keyte, J. W. & Waites, G. T. (1987) *J. Clin. Endocrinol. Metab.* **65**, 1067–1071.
15. Julkunen, M., Seppala, M. & Janne, O. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8845–8849.
16. Westwood, O. M. R., Chapman, M. G., Totty, N., Philip, R., Bolton, A. E. & Lazarus, N. R. (1988) *J. Reprod. Fertil.* **82**, 493–500.
17. Bell, S. C., Patel, S. R., Kirwan, P. H. & Drife, J. O. (1986) *J. Reprod. Fertil.* **77**, 221–231.
18. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
19. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
20. Gaye, P., Hue-Delahaie, D., Mercier, J. C., Soulier, S., Vilotte, J. L. & Furret, J. P. (1986) *Biochemie* **68**, 1097–1107.
21. Feinberg, A. P. & Vogelstein, B. (1982) *Anal. Biochem.* **132**, 6–13.
22. Sanger, F., Nicklen, S. & Coulson, A. D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
23. Nielsen, D. A. & Shapiro, D. J. (1986) *Nucleic Acids Res.* **14**, 5936.
24. Dobberstein, B., Garoff, H. & Warren, W. (1979) *Cell* **17**, 759–769.
25. Huhtala, M. L., Seppala, M., Narvanen, A., Palomaki, P., Julkunen, M. & Bohn, H. (1978) *Endocrinology* **120**, 2620–2622.
26. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
27. Shapiro, M. B. & Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
28. Padgett, R. A., Grabowski, P. J., Konarska, M. M. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.
29. Ali, S. & Clark, A. J. (1988) *J. Mol. Biol.* **199**, 415–426.
30. Harris, S., Ali, S., Anderson, S., Archibald, A. L. & Clark, A. J. (1988) *Nucleic Acids Res.* **16**, 10379–10380.
31. Ohshima, Y. & Gotoh, Y. (1987) *J. Mol. Biol.* **195**, 243–259.
32. Eperon, L. P., Estibeiro, J. P. & Eperon, I. C. (1986) *Nature (London)* **324**, 280–282.
33. Lear, A. L., Eperon, L. P., Wheatley, I. M. & Eperon, I. C. (1990) *J. Mol. Biol.* **211**, 103–115.
34. Reed, R. & Maniatis, T. (1988) *Cell* **46**, 681–690.
35. Nelson, K. K. & Green, M. R. (1988) *Genes Dev.* **2**, 319–329.
36. Robberson, B. L., Cote, G. J. & Berget, S. M. (1990) *Mol. Cell. Biol.* **10**, 84–94.
37. Cunningham, S., Else, A. J., Potter, B. V. L. & Eperon, I. C. (1991) *J. Mol. Biol.* **217**, 265–281.
38. Reed, R. & Maniatis, T. (1988) *Genes Dev.* **2**, 1268–1276.
39. Zhuang, Y., Goldstein, A. M. & Weiner, A. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2752–2756.
40. Noble, J. C. S., Prives, C. & Manley, J. L. (1988) *Genes Dev.* **2**, 1460–1475.
41. Smith, C. W. J., Porro, E. B., Patton, J. G. & Nadal-Ginard, B. (1989) *Nature (London)* **342**, 243–247.
42. Reed, R. (1989) *Genes Dev.* **3**, 2113–2123.
43. Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E. & Francke, U. (1987) *Nucleic Acids Res.* **15**, 10373–10391.
44. Jamieson, A. C., Vandeyar, M. A., Kang, Y. C., Kinsella, J. E. & Batt, C. A. (1987) *Gene* **61**, 85–96.
45. Shaw, G. & Kamen, R. (1986) *Cell* **46**, 659–667.
46. Bernstein, P. & Ross, J. (1989) *Trends Biochem. Sci.* **14**, 373–377.