Structure and expression of the guinea-pig α -lactalbumin gene

Jane E. LAIRD,*† Lucinda JACK,*§ Len HALL,†¶ Anthony P. BOULTON,* David PARKER,*|| and Roger K. CRAIG*

*Medical Molecular Biology Unit, Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, London W1P 7PN, U.K., and †Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, U.K.

The entire guinea-pig α -lactalbumin gene was isolated from a genomic DNA library constructed in the bacteriophage λ L47. The complete nucleotide sequence of the gene and its immediate 5' and 3' flanking sequences were determined and compared with those of the human and rat α -lactalbumin genes. This demonstrates that the size, organization and sequence of the exons is highly conserved between species, and reveals the presence of the highly conserved potential regulatory 'milk box' consensus sequence [RGAAGRAAA(N)TGGACAGAAATCAA(CG)TTTCTA] between positions –140 and –110 relative to the transcriptional start point. A guinea-pig LINE repeat sequence was also present in the 5' flanking region between positions –520 and –1195. Transfection of the α -lactalbumin gene cloned in a bovine papilloma virus vector into the mouse C127 and human MCF-7 mammary tumour cell-lines gave rise to stable but seemingly constitutive expression of α -lactalbumin. Expression was from the correct transcriptional start point, resulting in the accumulation of correctly processed mRNA and the secretion of α -lactalbumin into the culture medium.

INTRODUCTION

The synthesis of milk proteins, the major products of the epithelial cells of the lactating mammary gland, is regulated by an intricate combination of steroid and peptide hormones. Our current understanding of the manner in which hormones regulate milk-protein gene expression stems from studies on the whole animal, mammary explants in organ culture and epithelial cells cultured on collagen substrata, with the use of antibodies and cloned cDNA hybridization probes as a means of quantifying milk-protein gene expression. Such studies have identified that glucocorticoids and the peptide hormones insulin and prolactin are required for the synthesis of milk proteins (see Banerjee, 1976), and provide evidence that prolactin acts post-transcriptionally, resulting in the preferential accumulation of milk-protein mRNA in the cytoplasm (Houdebine *et al.*, 1978; Guyette *et al.*, 1979; Bathurst *et al.*, 1980). However, as yet, little is known about the intracellular mechanisms by which hormones, in particular the peptide hormone prolactin, modulate milk-protein gene expression, or the mechanisms that determine the tissue-specificity of milk-protein gene expression. To obtain such information requires the isolation, manipulation and expression of milk-protein genes in hormonally responsive cell-lines or transgenic animals.

Guinea-pig milk contains four major organ-specific proteins, the whey protein α -lactalbumin (Brew, 1972) and three caseins (Craig *et al.*, 1978). The mRNAs encoding all four guinea-pig milk proteins have been cloned (Craig *et al.*, 1981), and the complete nucleotide

sequences and deduced amino acid sequences determined for α -lactalbumin, casein A (α_{s2}) and casein B (α_{s1}) (Hall *et al.*, 1982, 1984a,b). In addition, parallel studies with sequence-specific cDNA probes demonstrate that guinea-pig α -lactalbumin and casein genes are regulated differentially during the onset of lactation (Burditt *et al.*, 1981).

In the present paper we describe the complete nucleotide sequence of the guinea-pig α -lactalbumin gene and its 5' and 3' flanking sequences. Comparison of this with the published sequences of the human and rat α -lactalbumin structural genes (see Qasba & Safaya, 1984; Hall *et al.*, 1987) shows an identical structural organization of four exons, identifies extensive homology within the 5' flanking region of all three genes, and the presence within this region, between positions –110 and –140 from the transcriptional start site, of the consensus sequence [RGAAGRAAA(N)TGGACAGAAATCAA(CG)TTTCTA] that we have previously reported to be present at the same position in human and rat α -lactalbumin, and five casein gene sequences (Hall *et al.*, 1987; see also Yu-Lee *et al.*, 1986). A guinea-pig LINE repeat sequence was also identified between positions –520 and –1195 within the 5' flanking region.

As a preliminary step towards determining the biological role of the candidate control element in the hormonal regulation or mammary specific expression of milk-protein genes, we demonstrate the stable but seemingly constitutive expression of the guinea-pig α -lactalbumin gene cloned into a bovine papilloma virus (BPV) vector, and transfected into the C127 mouse mammary tumour cell-line and MCF-7 human mammary tumour cell-line.

Abbreviation used: BPV, bovine papilloma virus.

† Present address: Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, U.K.

|| Present address: Wellcome Research Laboratories, Beckenham, Kent BR3 3BS, U.K.

§ Present address: Department of Animal Sciences, University of Maryland, Baltimore, MD 20742, U.S.A.

¶ To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00726.

MATERIALS AND METHODS

Materials

Restriction endonucleases, enzymes for DNA modification and synthesis and 'nick-translation' kits were obtained from Amersham International, Amersham, Bucks., U.K. [α - 32 P]dCTP (> 400 Ci/mmol), [α - 32 P]dATP (> 800 Ci/mmol) and GeneScreen *Plus* hybridization transfer membranes were obtained from Du Pont (U.K.) Ltd., Southampton, U.K. Specific oligonucleotides for sequence determination and primer extension were synthesized on a Du Pont Coder 300 automatic DNA synthesizer. All other chemicals and enzymes were of the highest grades available and obtained from sources previously described (see Craig *et al.*, 1976; Hall *et al.*, 1982, 1984*a,b*; Moore *et al.*, 1985). All tissue-culture media, plastics and the antibiotic G418 were obtained from GIBCO BRL, Paisley, Strathclyde, Scotland, U.K. The C127 cell-line (Lowy *et al.*, 1978) was a gift from Dr. C. C. Hentschel, Celltech Ltd., Slough, Berks., U.K., and the MCF-7 cell-line was obtained from Dr. C. McGrath, Michigan Cancer Foundation, Detroit, MI, U.S.A. The BPV vector pCGBPV9AB5 was a gift from Dr. P. Matthias, German Cancer Research Centre, Heidelberg, West Germany. All routine recloning procedures were as described elsewhere (see Maniatis *et al.*, 1982).

Construction and screening of a genomic library

A guinea-pig genomic DNA library was constructed in the bacteriophage λ replacement vector L47.1 (Loenen & Brammar, 1980) essentially as described by Hall *et al.* (1987). Briefly, lactating guinea-pig mammary gland DNA of high M_r was isolated from frozen tissue, and partially restricted with the enzyme *Sau*3A. Fragments of size 13–18 kb were selected by sucrose-density-gradient centrifugation, ligated into *Bam*HI-restricted bacteriophage λ L47.1 DNA arms, then packaged *in vitro* (Jeffreys *et al.*, 1980) by using extracts from the mutant lysogen strains BHB2690 and BHB2688. The resulting library (1.6×10^6 plaque-forming units) was titred on the *Escherichia coli* P2 lysogen Q359. A single recombinant bacteriophage containing sequences complementary to guinea-pig α -lactalbumin mRNA was identified by screening approximately 10^6 independent plaques with a high-specific-radioactivity 32 P-labelled guinea-pig α -lactalbumin cDNA probe (Hall *et al.*, 1982) essentially as described by Benton & Davis (1977). Subsequent restriction mapping and recloning of the α -lactalbumin gene into plasmid vectors followed established procedures (see Maniatis *et al.*, 1982).

Cell culture, transfection and selection

C127 cells were routinely maintained in Dulbecco's Minimal Essential Medium containing 15% (v/v) foetal calf serum and MCF-7 cells were maintained in a 1:1 (v/v) mixture of Eagle's Minimal Essential Medium and RPMI-1640 medium supplemented with 0.1% D-glucose, 1% (v/v) of a solution of non-essential amino acids, 1 mM-sodium pyruvate and 10% (v/v) foetal calf serum. All cells were cultured on plastic at 37 °C in an atmosphere of 5% CO₂ in air.

Cells were transfected with recombinant plasmid pLJ-1 by using the calcium phosphate technique (Graham & van der Eb, 1973) as modified by Southern & Berg (1982), with the inclusion of a glycerol shock.

Briefly, supercoiled pLJ-1 DNA (1 μ g) and carrier calf thymus DNA (25 μ g) were introduced into 5×10^5 cells in a 60 cm² flask. After 24 h exposure to DNA, the cells were trypsinized and replated at a 1:20 dilution. After a further 16 h, G418 (1.2 mg/ml) was added to the medium. Medium containing G418 was changed thereafter at 5–6-day intervals. Resistant colonies were trypsinized and resulting stable cell-lines were grown and maintained in selective medium.

Southern-blot analysis

Guinea-pig liver genomic DNA (20 μ g) was digested with the appropriate restriction enzymes, and the products were separated by electrophoresis on 0.8% (w/v) agarose gels and blotted on to NEN GeneScreen *Plus* membranes as described by Edbrooke *et al.* (1985). 32 P-labelled DNA hybridization probes (2×10^8 c.p.m./ μ g) were prepared by nick-translation (Rigby *et al.*, 1977). Bands were detected by autoradiography at –70 °C with Kodak X-Omat S X-ray film and Kodak Lanex intensifying screens.

DNA sequence analysis

Suitable restriction fragments were recloned in M13 mp8 or M13 mp9 vectors and DNA sequence analysis was performed by using the dideoxy chain termination method of Sanger *et al.* (1977, 1980). Sequencing was on both strands and across all restriction sites.

Isolation of RNA and analysis of α -lactalbumin RNA transcripts

Total RNA was isolated from tissue-culture cells or lactating mammary-gland tissue by using the guanidinium/CsCl method (Chirgwin *et al.*, 1979) as described by Maniatis *et al.* (1982) except that the tissue culture cells were lysed in 50 mM-Tris/HCl buffer, pH 7.6, containing 4 M-guanidinium thiocyanate, 10 mM-EDTA, 0.14 M-2-mercaptoethanol and 2% (w/v) sodium *N*-lauroylsarcosinate, and the viscosity was decreased by three 30 s bursts in a Polytron homogenizer, before centrifugation through a 5.7 M-CsCl cushion containing 100 mM-EDTA. Poly(A)-containing RNA was isolated from the recovered RNA as described by Craig *et al.* (1976). S₁-nuclease mapping of α -lactalbumin RNA transcripts was performed as described by Edbrooke *et al.* (1985) with T4 polynucleotide kinase 32 P-labelled DNA probes. Size determination of full-length α -lactalbumin mRNA transcripts was performed by RNA blotting on denaturing gels as described by Taylor *et al.* (1984).

Determination of transcriptional start site

The α -lactalbumin transcriptional start site was determined by primer extension as follows. An oligonucleotide primer (5'-AACAGGATGCCCCACCAG-3') was synthesized, complementary to nucleotide sequence encoding part of the signal peptide (amino acid residues –11 to –6) of the α -lactalbumin gene. A 10 ng portion of this primer was added to total poly(A)-containing RNA (0.25–1.5 μ g) in 10 μ l of 50 mM-Tris/HCl buffer, pH 8.3, containing 50 mM-NaCl and 9 mM-MgCl₂, and heated at 100 °C for 1 min to minimize secondary structure. It was then incubated at 65 °C for 10 min and finally allowed to cool slowly to 37 °C in a beaker of hot water, over a period of about 2 h. The volume was increased to 25 μ l by the addition of suitable stock solutions to achieve a

final composition of 50 mM-Tris/HCl buffer, pH 8.3, 3.6 mM-MgCl₂, 20 mM-NaCl, 5 mM-dithiothreitol, 50 μ g of actinomycin D/ml, 500 μ M-dCTP, 500 μ M-dGTP, 500 μ M-dTTP and 15 μ M-[α -³²P]dATP (50 μ Ci). Rous-associated-virus-2 reverse transcriptase (10 units) was added and incubated at 37 °C for 15 min. After addition of 1.25 μ l of a 10 mM-dATP 'chase' solution, incubation was continued at 37 °C for a further 15 min. The reaction was then terminated by the addition of 150 μ l of 25 mM-EDTA solution, pH 7, and the extended primer-template was concentrated by ethanol precipitation in the presence of 5 μ g of *E. coli* tRNA. Finally, the extended primer was redissolved in 10 μ l of 80% (v/v) formamide containing 1 mM-EDTA, 10 mM-NaOH and 10 mg of Bromophenol Blue/ml, and samples were analysed on standard 6% (w/v) polyacrylamide/urea sequencing gels.

Identification of α -lactalbumin in tissue-culture medium by immunoblotting

The isolation of α -lactalbumin secreted into tissue-culture medium was based on its solubility in acid ethanol (Brew & Campbell, 1967). Medium (1.2 ml) harvested from pLJ-1 transfected and non-transfected cells was centrifuged in a Microfuge to remove debris, and 1.0 ml of the resulting supernatant was freeze-dried. The residue was extracted three times with 0.03 M-HCl in 85% (v/v) ethanol (200 μ l) and the pooled extracts were dried *in vacuo*. The resulting residue was then dissolved in water (200 μ l) and 1 μ g of bovine serum albumin was added as carrier. Protein was then precipitated by the addition of trichloroacetic acid to a final concentration of 12.5% (w/v), the precipitate was recovered by centrifugation and the supernatant was discarded. The presence of α -lactalbumin in the precipitated material was then determined by SDS/polyacrylamide-gel electrophoresis and Western blotting as described by Moore *et al.* (1985) with a polyclonal rabbit antiserum raised against guinea-pig α -lactalbumin (see Craig *et al.*, 1976).

RESULTS AND DISCUSSION

Structure of the guinea-pig α -lactalbumin gene

A guinea-pig genomic DNA library was constructed in bacteriophage λ L47.1 (10⁶ independent plaques) and screened with a ³²P-labelled guinea-pig α -lactalbumin

cDNA fragment derived from plasmid pgpK-9 (see Craig *et al.*, 1981; Hall *et al.*, 1982). A number of signals were detected on the primary screen, and one designated L ϕ 4GI was successfully plaque-purified through four subsequent cycles of screening.

Restriction analysis and mapping of L ϕ 4GI showed that the recombinant bacteriophage contained 15 kb of guinea-pig genomic DNA, comprising the entire α -lactalbumin gene and about 11 kb of upstream sequence (Laird, 1985). To facilitate mapping and subsequent nucleotide sequence analysis, a 5.5 kb *Hind*III fragment containing a fragment of vector DNA at the 3' end was subcloned into plasmid pAT153 (designated pgpH68), and the entire nucleotide sequence of the α -lactalbumin gene and its immediate flanking regions (*Hind*III–*Eco*RI fragment) was determined on both strands and across all restriction sites by further subcloning into bacteriophage M13, by using the strategy outlined in Fig. 1.

Comparison of the resulting nucleotide sequence (Fig. 2) with that of guinea-pig α -lactalbumin cDNA (Hall *et al.*, 1982) revealed the presence of four exons, and demonstrated two disparities between sequences common to genomic and cloned cDNA sequences. Firstly, in exon 4 the termination codon (TGA) occurred at the expected position immediately after Gln-123. This contrasted with one out of six α -lactalbumin cDNA clones previously analysed, which, as the result of a T $\rightarrow$ C base change at this position, encoded an α -lactalbumin molecule with a 36-amino-acid-residue C-terminal extension before the first in-phase termination codon. This we now presume is a reflection of a cloning artifact (see Hall *et al.*, 1982). Secondly, homology between the 5' untranslated region of the cloned α -lactalbumin cDNA and the corresponding genomic sequence was limited to the first 28 nucleotide residues upstream from the initiating methionine in exon 1. The remaining 56 nucleotide residues present at the 5' end of the cloned α -lactalbumin cDNA sequence (see Hall *et al.*, 1982) proved complementary to a sequence present in the 3' untranslated region of the mRNA, and therefore represents a cloning artifact (Laird, 1985).

The positions of the exon-intron boundaries in the guinea-pig α -lactalbumin gene are similar to those in the corresponding human and rat genes, occurring within the codons for Trp-26 (intron I), Lys-79 (intron II) and

Fig. 1. Genomic organization and sequencing strategy for the guinea-pig α -lactalbumin gene

Sites for enzymes utilized for DNA sequencing of the guinea-pig α -lactalbumin gene cloned into pgpH68 are shown. Arrows indicate the direction and extent of individual sequencing runs. Open circles (O) identify sequences obtained by using internal oligonucleotide primers.

1988

Trp-104 (intron III). All three residues are conserved between rat and guinea pig, whereas Trp-26 is replaced by Leu-26 in the human protein. As discussed previously (Hall *et al.*, 1987), conservation of exon-intron organization indicates that this was established before the divergence of the species. The size of the introns was also generally conserved between species, though this was dependent on the presence or otherwise of inserted repetitive elements. Thus introns I and II are of similar size in guinea-pig and rat α -lactalbumin genes, and intron II and III are of similar size in human and guinea pig. The disparities between human and rat reflect the presence of repetitive elements (see Hall *et al.*, 1987).

Analysis of the nucleotide sequence of the guinea-pig α -lactalbumin gene for conserved sequences of potential regulatory or functional significance revealed a number of features in common with other mammalian genes, and some peculiar to previously determined α -lactalbumin and casein genomic sequences. We have established, by primer extension of a synthetic oligonucleotide annealed to lactating guinea-pig mammary-gland total poly(A)-containing RNA, that transcription (+1) was initiated with a uracil residue 29 bases from the initiating methionine codon in exon 1 (see Fig. 3). This was preceded by an A+T-rich region TAATAAAA (centred at -27), found within the promoter region of most eukaryotic genes. However, in common with other α -lactalbumin genes so far examined, the so-called CAAT box normally found at a position between -80 and -70 was not present. However, the milk-protein consensus sequence, the 'milk box' [RGAAGRAAA(N)TGGAC-AGAAATCAA(CG)TTTCTA], previously identified between -140 and -110 in α -lactalbumin (Hall *et al.*, 1987) and casein genomic sequences (Yu-Lee *et al.*, 1986) was also present at an identical position in the guinea-pig α -lactalbumin gene (see Figs. 2 and 4), though the degree of homology relative to the consensus sequence was 73.1% when compared with 72.4% and 75.8% for rat and human α -lactalbumin genes respectively. Generally, as emphasized by DIAGON analysis (Staden, 1982) of human and guinea-pig α -lactalbumin genes and direct sequence comparisons (results not shown), there was significant sequence homology within the 5' flanking regions until position -520, at which point the guinea-pig sequence diverges, the point of insertion of a guinea-pig LINE repeat sequence (see below).

Within the structural gene the dinucleotide splice consensus signals GT and AG occur at the 5' and 3' ends respectively of the introns (Mount, 1982), while at the 3' end cleavage and polyadenylation occur at a site G(A) 14 residues downstream from the '3'-cleavage' signal (AATAAA). The cleavage site immediately precedes a T+G-rich region containing a (T)-rich sequence that has also been implicated in precursor cleavage and polyadenylation in a number of other eukaryotic genes (see Birnstiel *et al.*, 1985).

Identification of a guinea-pig LINE sequence within the 5' flanking region of the guinea-pig α -lactalbumin gene

Southern-blot analysis of guinea-pig genomic DNA

Fig. 3. Mapping of the transcriptional start point of α -lactalbumin RNA by primer extension

Primer extension of α -lactalbumin RNA transcripts in a number of RNA preparations was performed as described in the Materials and methods section, and the products were resolved by electrophoresis on a 6% (w/v) urea/polyacrylamide sequencing gel. A sequencing ladder was performed in parallel by using the same oligonucleotide primer and an M13 template that contained an *SspI*/*NaeI* fragment comprising exon 1 and 5' flanking regions of the gene (see Fig. 1). (a) (i) Primer extension of lactating guinea-pig mammary-gland poly(A)-containing RNA (0.08 μ g); (ii) DNA sequencing ladder electrophoresed in parallel. (b) Primer extension of (i) lactating guinea-pig mammary gland poly(A)-containing RNA, and total cellular poly(A)-containing RNA isolated from (ii) C127 cells transfected with pLJ-1 (0.5 μ g); (iii) MCF-7 cells transfected with pLJ-1 (0.5 μ g); (iv) non-transfected C127 cells (1.5 μ g); (v) non-transfected MCF-7 cells (1.5 μ g).

Fig. 2. Complete DNA sequence of the guinea-pig α -lactalbumin gene and immediate flanking sequences

The sequence is numbered from the transcriptional start site (+1; see Fig. 3). The site of polyadenylation in the transcript is indicated, and the 'milk box' (-110 to -140) has been underlined.

Overall consensus:	R G A A G R A A A(N) T G G A C A G A A A - T C A A (C G) T T T C T A	Homology (%)
Guinea-pig α -lactalbumin:	 G C . . C A G T T G - - -	73.1
Human α -lactalbumin:	 G C G C - - - G C G G	75.8
Rat α -lactalbumin:	. . . G G T G C - - - G G C G	72.4
Rat α -casein:	 T - A - T . . - - C T	86.2
Rat β -casein:	. A - . A . . . A . . . - C . . T - -	82.7
Rat γ -casein:	 C - C C . G G . . - -	79.3
Bovine α_{s1} -casein:	. . . G . . . C T - - . A C A . T G . . - - . . . C T	69.0
Guinea-pig α_{s2} -casein:	 C - G - - - -	79.3

Fig. 4. Conserved region within the 5' flanking sequence of eight milk-protein genes

Sequences from about -140 to -110 of three α -lactalbumin and five casein genes were aligned with the introduction of gaps (-) to maximize homology. Dots represent homology with consensus, and percentage homology for each sequence is expressed relative to the consensus sequence.

with a guinea-pig α -lactalbumin cDNA probe (pgpK-9) showed, with the use of a number of enzyme digests, a banding pattern consistent with the presence of a single unique α -lactalbumin gene in the guinea-pig genome and the absence of other closely related genes (Fig. 5a). However, parallel analysis with the whole of the 3.5 kb genomic fragment cloned in pgpH68 as the hybridization probe revealed a complex banding pattern not observed with an α -lactalbumin cDNA probe, consistent with the presence of a number of repetitive sequences within intron or flanking regions of the gene (results not shown). Subsequent re-analysis of the same blot with the *NaeI*-*HindIII* fragment of pgpH68 representative of the entire 5' flanking region and part of exon 1 of the guinea-pig gene showed (Fig. 5b) that a repeat sequence present in this fragment hybridized strongly to a 1.9 kb *Bgl*II guinea-pig genomic repetitive sequence. Comparison of sequences present between nucleotides -520 and -1195 of the guinea-pig gene with other known repetitive sequence revealed a high degree of homology (Table 1) with the mammalian long interspersed repeat families (LINES).

Alignment of the region of homology between the guinea-pig repetitive sequence and the primate and rodent LINE sequences indicate that (i) the guinea-pig LINE sequence adjacent to the α -lactalbumin gene is a truncated copy, (ii) the sequence is homologous to a region near the 3' end of the 'full-length' LINE, and (iii) within the limited guinea-pig sequence available the extent of homology with the primate and rodent L1 families is very similar (see Table 1). LINE sequences have previously been shown to be associated with many scattered chromosome locations; they flank genes, reside within introns and are associated with centromeric satellite DNA (for a review see Singer & Skowronski, 1985).

Expression of the guinea-pig α -lactalbumin gene

Construction of BPV vectors and transfection into cell-lines of mammary origin. The isolation and sequence analysis of the guinea-pig α -lactalbumin gene provides structural details but does not define a functional gene, in spite of the presence of the expected conserved regulatory sequences invariably necessary for transcription and mRNA maturation. To determine whether the sequences cloned in pgpH68 represent a functional gene, we have subcloned, using *Bgl*II linkers, the *HindIII*-*Bgl*II fragment of pgpH68, which contains the complete structural gene and flanking sequences, into a unique *Bam*HI site in the BPV-based shuttle vector pCGBPV9 Δ B5, creating a plasmid designated pLJ-1 (see Fig. 6a). Restriction

analysis of four separate isolates all showed the α -lactalbumin gene inserted in the same orientation. The BPV vector is a derivative of pCGBPV9, which contains a G418 selectable marker (see Matthias *et al.*, 1983) from which the second *Bam*HI site adjacent to the BPV enhancer had been deleted (P. D. Matthias, personal communication).

Four cell-lines were established from G418 resistant foci of pLJ-1 transfected with the use of calcium phosphate into mouse C127 cells, a cell-line of mammary origin, and a single cell-line from pLJ-1-transfected

Fig. 5. Identification of repeat sequences within the 5' flanking region of the guinea-pig α -lactalbumin gene

Guinea-pig genomic DNA (20 μ g per track) was restricted with a number of different enzymes, and the fragments were separated by electrophoresis and blotted on to membranes as described in the Materials and methods section. The membrane was then probed with (a) 32 P-labelled, nick-translated α -lactalbumin cDNA (pgpK-9), or (b) a 5' flanking region probe of pgpH-68 (*HindIII*-*NaeI*). Tracks (i), (ii), (iii) and (iv) contain genomic DNA restricted with the enzymes *Bgl*II, *Eco*RI, *HindIII* and *Sst*I respectively. *HindIII*-restricted bacteriophage λ DNA was electrophoresed in parallel as size markers. Filters were exposed to X-ray film for (a) 48 h and (b) 3 h.

Table 1. Homology between the guinea-pig LINE sequence and those of other species

The guinea-pig repetitive sequence associated with the α -lactalbumin gene was compared with various LINE sequences present in the EMBL nucleotide sequence database, by use of the GAP alignment program (Devereux *et al.*, 1984). Below is shown the alignment of the guinea-pig partial LINE1 sequence with the 'full-length' primate and rodent LINE1 sequences. The long open reading frames (ORFs) found in some of the primate LINE sequences are indicated. Regions of homology identified by sequence analysis are indicated by heavy hatching and regions of homology established by cross-hybridization are shown in light hatching.

Repetitive sequence	Homology (%)	EMBL ID	Reference
Human LI (KpnI)	64	HSKPNI03	Hattori <i>et al.</i> (1985)
Human LI mRNA	65	HSRSKP08	DiGiovanni <i>et al.</i> (1983)
Monkey LI (KpnI)	62	MOKPNRSA	Lerman <i>et al.</i> (1983)
Rat LI	64, 62	RNRL13, RNRL115	Soares <i>et al.</i> (1985)
Mouse LI	53	MMDRF3	Gebhard & Zachau (1983)
Mouse MIF repeat	64, 63	MMRSMF35, MMRSMF37	Brown & Piechaczyk (1983)

MCF-7 cells, a cell-line of human mammary tumour origin. Transfected and parental cell-lines were grown on plastic, then harvested, and the cells were used as a source of genomic DNA and total cellular RNA. Culture medium surrounding the cells was also retained. Southern-blot analysis of restricted genomic DNA confirmed the presence of the guinea-pig α -lactalbumin gene in three of the C127 cell-lines (clones 1, 2 and 3) and the MCF-7 cell-line at a copy number of approximately one to ten copies/cell. In the remaining C127-derived cell-line (clone 4) the α -lactalbumin gene was present at a copy number of about 100 copies/cell. There was no evidence for the rearrangement of the guinea-pig α -lactalbumin gene in any of the cell-lines, and Southern-blot analysis of BPV sequences indicated that the plasmid remained episomal and had not integrated into the genome after transfection (Jack, 1986).

α -Lactalbumin gene expression in transfected cell-lines.

Evidence for correct expression of the guinea-pig gene was obtained by a combination of RNA blotting to determine mRNA size, S_1 -nuclease mapping to determine the nature of the 3' end of RNA transcripts, and primer extension to determine whether transcription was occurring from the normal transcriptional start point. In addition, culture medium was analysed by Western blotting with affinity-purified antisera raised against guinea-pig α -lactalbumin, to determine the presence or otherwise of the secreted mature protein in the culture medium. Lactating guinea-pig mammary-gland RNA or α -lactalbumin was included as appropriate in all experiments as a control, and as a means of estimating the relative amounts of α -lactalbumin RNA transcripts and of secreted α -lactalbumin. In the first instance all the C127- and MCF-7-derived cell-lines were examined for α -lactalbumin gene expression. All expressed α -lactalbumin transcripts as judged by S_1 -nuclease mapping (see below),

but one (C127 clone 3) accumulated α -lactalbumin RNA transcripts at a 10-fold greater concentration than did clones 1 and 2 or the transfected MCF-7 cell-line. C127 clone 4, which had the highest copy number, showed the lowest concentration of α -lactalbumin transcripts (Jack, 1986). All presented data compare guinea-pig α -lactalbumin gene expression in C127 cells (clone 3), MCF-7 cells and the lactating guinea-pig mammary gland.

Analysis of α -lactalbumin expression in C127 (clone 3) cells by RNA blotting of total cellular RNA (Fig. 6b) showed the presence of a transcript of size identical with that of guinea-pig α -lactalbumin mRNA, when hybridized with an α -lactalbumin cDNA probe. The transcript size (750 nucleotide residues) and abundance were consistent with the constitutive expression of a correctly processed α -lactalbumin mRNA at a level of synthesis of about 5% that of the lactating mammary gland, or about 1000 copies per cell. No α -lactalbumin RNA transcripts were present in the C127 control cell-line. Parallel analysis of transfected MCF-7 cells identified a band of the same size, but only after prolonged exposure, indicating a very much lower level of expression (L. Jack, unpublished work).

The presence of α -lactalbumin transcripts with correctly processed 3' ends in transfected C127 and MCF-7 cell-lines was then demonstrated by S_1 -nuclease mapping with a *Bam*HI-*Bgl*II DNA fragment comprising sequence from exon 4 and 3' flanking sequences. This produced the expected 147-nucleotide-residue protected fragment with RNA isolated from the lactating mammary gland, transfected C127 cells and also, but at a very much lower level, transfected MCF-7 cells (Fig. 6c). No protection was observed with RNA from control C127 and MCF-7 cells. An additional protected band of 137 nucleotide residues was also apparent with total RNA, and to a lesser extent poly(A)-containing RNA isolated from the

Fig. 6. Identification of correctly processed α -lactalbumin mRNA transcripts in pLJ-1 transfected cell-lines

(a) Plasmid pLJ-1 containing the guinea-pig α -lactalbumin gene (*Hind*III–*Bgl*II fragment) cloned into the *Bam*HI site of pCGBP9 Δ B5. The arrows indicate the direction of transcription. (b) Northern-blot analysis. Total RNA (20 μ g per track) isolated from (1) non-transfected C127 cells, from (2) C127 cells transfected with pLJ-1 and from (3) lactating guinea-pig mammary gland was analysed by RNA blotting, and probed with a ³²P-labelled nick-translated α -lactalbumin cDNA probe (pgp-K9) as described in the Materials and methods section. (c) S₁-nuclease mapping of the 3' end of α -lactalbumin mRNA. S₁-nuclease mapping with a ³²P-labelled *Bam*HI–*Bgl*II fragment isolated from pgpH-68 comprising exon 4 and 3' flanking sequence was performed on (1) total RNA (50 ng) and (2) poly(A)-containing RNA (10 ng) from the lactating guinea-pig mammary gland, on (3) total RNA (40 μ g) and (4) poly(A)-containing RNA (0.8 μ g) from C127 cells transfected with pLJ-1, and on (5) total RNA (40 μ g) from non-transfected C127 cells. In a separate series of experiments S₁-nuclease mapping was performed on (6) total RNA (20 μ g) from non-transfected MCF-7 cells, on (7) total RNA (20 μ g) from MCF-7 cells transfected with pLJ-1, and on (8) total RNA (5 ng) from the lactating guinea-pig mammary gland, all as described in the Materials and methods section. Size markers are a ³²P-labelled *Hin*fI digest of plasmid pAT153 DNA. Abbreviation: ntd, nucleotide residues.

lactating mammary gland and transfected C127 cells. This we presume is a component of the non-polyadenylated α -lactalbumin RNA, and may represent an intermediate in the turnover of α -lactalbumin mRNA. The absence of this 137-nucleotide-residue intermediate from transfected MCF-7 cells may simply reflect the very much lower concentration of accumulated α -lactalbumin mRNA in MCF-7 cells when compared with that of C127 (clone 3) cells.

Primer extension of total cellular poly(A)-containing RNA isolated from transfected C127 and MCF-7 cells, analysed in parallel with poly(A)-containing RNA from the lactating guinea-pig mammary gland (Fig. 3b), again confirmed the presence of α -lactalbumin RNA transcripts in the transfected, but not control, cells, and demonstrated that initiation of transcription was with a uracil residue (+1), the α -lactalbumin transcriptional start site used in the mammary gland. Taken *in toto*, a combination of S_1 -nuclease mapping, primer extension and RNA blotting demonstrates that the guinea-pig α -lactalbumin gene cloned and expressed in a BPV host vector system is transcribed from its own promoter, then correctly spliced and polyadenylated, in C127 or MCF-7 cells.

Examination of extracted cell-culture medium also confirmed the synthesis and secretion of processed α -lactalbumin protein by transfected C127 and MCF-7 cell lines, but not the untransfected controls (Fig. 7). In quantitative terms the levels of secretion in both cell-lines represented an accumulation of about 5–10 ng of α -

lactalbumin per ml of overnight culture, when compared with a guinea-pig α -lactalbumin standard. Given the disparity in α -lactalbumin mRNA concentrations in the transfected C127 and MCF-7 cell-lines, it was perhaps surprising that α -lactalbumin accumulation in the culture medium showed equivalence. Since recovery of α -lactalbumin from the culture medium is quantitative, we presume that this reflects greater turnover of α -lactalbumin within the secretory pathway of C127 cells relative to the MCF-7 cells, or more efficient secretion from MCF-7 cells.

Attempts to manipulate the extent of α -lactalbumin mRNA accumulation or protein secretion in the transfected MCF-7 cell-line by growth in the presence of various concentrations of insulin, cortisol or sheep prolactin failed to affect the relative accumulation of α -lactalbumin RNA or α -lactalbumin secretion, whether the cells were grown on plastic, collagen (Primeria flasks; Pittler *et al.*, 1985) or floating collagen gels (Emerman & Pitelka, 1977). Thus, in spite of the reported presence of insulin, glucocorticoid and prolactin receptors in this cell-line (Monaco & Lippman, 1977; Horwitz *et al.*, 1978; Shiu, 1979), only constitutive expression of the transfected guinea-pig α -lactalbumin gene was observed.

CONCLUSION

We have recently described the organization, sequence and chromosomal localization of the human α -lactalbumin gene (Davies *et al.*, 1987; Hall *et al.*, 1987). Our studies on the guinea-pig α -lactalbumin gene show many similarities. Both genes, on the basis of Southern-blotting at high stringency, would appear to be unique, though chromosomal localization studies on the human α -lactalbumin gene at lower stringency provide circumstantial evidence for a second related gene in the region 12p11, as compared with α -lactalbumin at 12q13 (see Davies *et al.*, 1987). The exon-intron organization of rat (Qasba & Safaya, 1984), human and guinea-pig genes is also identical, and variation in intron size can be accounted for by the presence of inserted repetitive elements. The presence of the 'milk box' consensus sequence between positions -110 and -140 of the 5' flanking sequence of rat, human and now guinea-pig α -lactalbumin genes and all published casein gene sequences (see Yu-Lee *et al.*, 1986), but its absence from the lysozyme gene (see Hall *et al.* 1987), add weight to mounting circumstantial evidence that the 'milk box' may play a regulatory role as a cis-acting sequence involved in either hormone regulation or tissue-specific expression of α -lactalbumin and casein genes in the mammary gland, through interaction with as yet unknown regulatory factors. In this respect our own preliminary evidence based on gel retardation analysis with mammary gland nuclear extracts would support this view (A. Maschio, unpublished work).

Studies on the expression of the guinea-pig α -lactalbumin gene by using a BPV-based expression vector in two cell-lines of mammary origin demonstrated unexpectedly high constitutive expression of the gene from its own promoter, correctly spliced and polyadenylated, with secretion of α -lactalbumin into the medium, but no evidence for modulation of gene expression by hormones. A number of other genes, including the rat preproinsulin gene, human β -globin, HLA genes and HBVs Ag genes are all correctly transcribed and constitutively expressed

Fig. 7. Identification of α -lactalbumin in the culture medium by immunoblotting

Immunoblotting with affinity-purified rabbit anti- α -lactalbumin IgG was performed as described in the Materials and methods section. α -Lactalbumin was acid/ethanol-extracted from 1 ml of cell medium, and the precipitated proteins were separated by SDS/polyacrylamide-gel electrophoresis before immuno-analysis: track 1, guinea-pig α -lactalbumin standard (100 ng); track 2, medium from non-transfected C127 cells; track 3, medium from C127 cells transfected with pLJ-1; track 4, medium from non-transfected MCF-7 cells; track 5, medium from MCF-7 cells transfected with pLJ-1.

in cells transformed by BPV hybrid plasmids. Constructs that use metallothionein or mouse mammary tumour virus long terminal repeat chimeras show increased mRNA accumulation on induction with Cd^{2+} or dexamethasone, and human β -interferon is produced constitutively and responds to induction by poly(I)·poly(C) (for review see Campo, 1985). The reason for the high level of constitutive expression, and the lack of response of the guinea-pig α -lactalbumin gene to hormone modulation, is not obvious, but may in part reflect the presence of the BPV enhancer downstream of the cloned gene (see also Sarver *et al.*, 1985). Recent experiments with a rat α -casein minigene cloned into a BPV vector and transfected into COMMA-1D cells, a mouse cell-line of mammary origin that expresses β -casein from its endogenous gene in a hormone-responsive manner, also showed constitutive but not hormonally regulated expression of the rat α -casein minigene (Yu-Lee & Rosen, 1988). Thus it would appear that although many aspects of milk-protein gene expression, for example protein turnover, post-translational modification and secretion, can be studied by the use of BPV-based vectors transfected into cell-lines of mammary origin, it may not be the vector of choice for identifying cis-acting sequences required for the tissue-specific expression and hormonal regulation of milk-protein genes. Other vectors and transgenic mouse models may be better suited to such studies.

We thank the Medical Research Council for supporting this work, and Dr. P. D. Matthias for the gift of the BPV vector.

REFERENCES

- Banerjee, M. R. (1976) *Int. Rev. Cytol.* **47**, 1–97
- Bathurst, I. C., Craig, R. K., Herries, D. G. & Campbell, P. N. (1980) *Eur. J. Biochem.* **109**, 183–191
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182
- Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359
- Brew, K. (1972) *Eur. J. Biochem.* **27**, 341–353
- Brew, K. & Campbell, P. N. (1967) *Biochem. J.* **102**, 258–264
- Brown, S. D. M. & Piechaczyk, M. (1983) *J. Mol. Biol.* **165**, 249–256
- Burditt, L. J., Parker, D., Craig, R. K., Getova, T. & Campbell, P. N. (1981) *Biochem. J.* **194**, 999–1006
- Campo, M. S. (1985) in *DNA Cloning* vol. 2: A Practical Approach (Glover, D. M., ed.), pp. 213–238, IRL Press, Oxford
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Craig, R. K., Brown, P. A., Harrison, O. S., McIlreavy, D. & Campbell, P. N. (1976) *Biochem. J.* **160**, 57–74
- Craig, R. K., McIlreavy, D. & Hall, R. L. (1978) *Biochem. J.* **173**, 633–641
- Craig, R. K., Hall, L., Parker, D. & Campbell, P. N. (1981) *Biochem. J.* **194**, 989–998
- Davies, M. S., West, L. F., Davis, M. B., Povey, S. & Craig, R. K. (1987) *Ann. Hum. Genet.* **51**, 183–188
- Dereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- DiGiovanni, L., Haynes, S. R., Misra, R. & Jelinek, W. R. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6533–6537
- Edbrooke, M. R., Parker, D., McVey, J. H., Riley, J. H., Sorenson, G. D., Pettergill, O. S. & Craig, R. K. (1985) *EMBO J.* **4**, 715–724
- Emerman, J. T. & Pitelka, D. R. (1977) *In Vitro* **13**, 316–328
- Gebhard, W. & Zachau, H. G. (1983) *J. Mol. Biol.* **170**, 255–270
- Graham, F. L. & van der Eb, A. J. (1973) *Virology* **52**, 456–461
- Guyette, W. A., Matusik, R. J. & Rosen, J. M. (1979) *Cell* **17**, 1013–1023
- Hall, L., Craig, R. K., Edbrooke, M. R. & Campbell, P. N. (1982) *Nucleic Acids Res.* **10**, 3503–3515
- Hall, L., Laird, J. E. & Craig, R. K. (1984a) *Biochem. J.* **222**, 561–570
- Hall, L., Laird, J. E., Pascall, J. C. & Craig, R. K. (1984b) *Eur. J. Biochem.* **138**, 585–589
- Hall, L., Emery, D. C., Davies, M. S., Parker, D. & Craig, R. K. (1987) *Biochem. J.* **242**, 735–742
- Hattori, M., Hidaka, S. & Sakaki, Y. (1985) *Nucleic Acids Res.* **13**, 7813–7827
- Horwitz, K. B., Zava, D. T., Thilager, A. K., Jensen, E. M. & McGuire, W. L. (1978) *Cancer Res.* **38**, 2434–2437
- Houdebine, L.-M., Devinoy, E. & Delouis, C. (1978) *Biochimie* **60**, 57–63
- Jack, L. J. (1986) Ph.D. Thesis, University of London
- Jeffreys, A. J., Wilson, V., Wood, D., Simons, J. P., Kay, R. M. & Williams, J. G. (1980) *Cell* **21**, 555–564
- Laird, J. E. (1985) Ph.D. Thesis, University of London
- Lerman, M. I., Thayer, R. E. & Singer, M. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3966–3970
- Loenen, W. A. M. & Brammar, W. J. (1980) *Gene* **10**, 249–259
- Lowy, D. R., Rand, E. & Scholnick, E. M. (1978) *J. Virol.* **26**, 291–298
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Matthias, P. D., Bernard, H. V., Scott, A., Brady, G., Hashimoto-Gotch, T. & Schutz, G. (1983) *EMBO J.* **2**, 1487–1492
- Monaco, M. E. & Lippman, M. E. (1977) *Endocrinology* (Baltimore) **101**, 1238–1246
- Moore, A., Boulton, A. P., Heid, H. W., Jarasch, E.-D. & Craig, R. K. (1985) *Eur. J. Biochem.* **152**, 729–737
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472
- Pittner, R. A., Fears, R. & Brindley, D. N. (1985) *Biochem. J.* **225**, 455–462
- Qasba, P. K. & Safaya, S. K. (1984) *Nature (London)* **308**, 377–380
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Sarver, N., Muschel, R., Byrne, J. C., Khoury, G. & Howley, P. M. (1985) *Mol. Cell. Biol.* **5**, 3507–3516
- Shiu, R. P. C. (1979) *Cancer Res.* **39**, 4381–4386
- Singer, M. F. & Skowronski, J. (1985) *Trends Biochem. Sci.* **10**, 119–122
- Soares, M. B., Schon, E. & Efstratiadis, A. (1985) *J. Mol. Evol.* **22**, 117–133
- Southern, P. J. & Berg, P. (1982) *J. Appl. Genet.* **1**, 327–341
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961
- Taylor, J. B., Craig, R. K., Beale, D. & Ketterer, B. (1984) *Biochem. J.* **219**, 223–231
- Yu-Lee, L.-Y. & Rosen, J. M. (1988) *Mol. Endocrinol.*, in the press
- Yu-Lee, L.-Y., Richter-Mann, L., Couch, C. H., Stewart, A. F., Mackinlay, A. G. & Rosen, J. M. (1986) *Nucleic Acids Res.* **14**, 1883–1902