The construction, identification and characterisation of plasmids containing human α-lactalbumin cDNA sequences

Len Hall, Michael S. Davies and Roger K. Craig* Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, London, W1P 7PN, UK

Received 22 October 1980

ABSTRACT

We describe the cloning of double-stranded cDNA synthesized from lactating human mammary gland total poly(A)-containing RNA, into the EcoRI site of the plasmid pAT153. Nine recombinants were shown to contain α-lactalbumin cDNA sequences as determined by positive hybridisation translation of complementary RNA. Restriction enzyme maps were determined for six of these. Alignment of the restriction map with the known amino acid sequence of human α-lactalbumin provided evidence that two plasmids, designated phO-53 and phB-35, contained the complete coding sequence of the primary translation product (pre-α-lactalbumin). Hybridisation studies using purified human, monkey and guinea-pig α-lactalbumin cDNA demonstrated that greater nucleotide sequence divergence has occurred within the rodents than the primates, and that rodent α-lactalbumin mRNAs retain regions of homology with primate α-lactalbumin mRNAs.

INTRODUCTION

In the Western World breast cancer accounts for as much as 10% of all fatalities in the female population. Since one third of all metastatic breast carcinomas show regression in response to hormone therapy, there has been a need to predict which tumours are of the hormone-dependent type (1,2). Recently it has been suggested that the presence in tumours of oestrogen receptors and milk proteins, may indicate an intact oestrogen-receptor mechanism, and may therefore provide a useful marker for the identification of tumours responsive to hormone therapy (3,4). Unfortunately reports regarding the presence or otherwise of milk proteins in tumour tissue or serum have proved inconclusive, possibly attributable to problems associated with the radioimmunoassay invariably used for such studies (5,6). In addition to technical problems, the immunological approach...
only monitors the final product of hormone action, the mature protein, and not the initial RNA transcripts. A better approach would seem to be an assay designed to detect differences in transcriptional rather than translational activity.

As a preliminary step towards this objective, we recently described the isolation of poly(A)-containing RNA from normal lactating human mammary gland, and demonstrated that this preparation directs the synthesis of casein and α-lactalbumin in vitro (7).

Here we describe the construction of a library of cDNA plasmids from this RNA preparation, and the isolation and characterisation of plasmids containing α-lactalbumin cDNA sequences. We also demonstrate, using α-lactalbumin specific cDNA hybridisation probes, that greater nucleotide sequence divergence has occurred within the rodent species studied than within the primates studied. We report elsewhere on similar hybridisation studies designed to determine the presence or otherwise of α-lactalbumin RNA transcripts in human breast tumour tissues, and human epithelial cell lines of mammary origin (8).

MATERIALS AND METHODS
Isolation of total cellular poly(A)-containing RNA from human mammary gland

Total cellular poly(A)-containing RNA was isolated from a large (2 g) retention cyst, obtained from a nursing mother 15 days postpartum. This RNA preparation had previously been shown to direct the synthesis of substantial levels of α-lactalbumin and caseins when translated in a wheat germ cell-free protein synthesizing system (7). Residual DNA was removed from the preparation by incubation with 20 μg/ml iodoacetylated deoxyribonuclease (9) for 15 min at 4°C in 10 mM NaCl, 10 mM magnesium acetate, 10 mM Tris-Cl, pH 7.4, containing 200 μg/ml E.coli tRNA.

Synthesis of dT-tailed double-stranded cDNA

The preparation of 3H-labelled cDNA (1440 cpm/ng) complementary to poly(A)-containing RNA, synthesis of the second DNA strand, cleavage of the "hairpin-loop", size selection of
the resulting double-stranded cDNA on agarose gels, and 
extension of the 3'-termini with dT-residues were as described 
elsewhere for the construction of a cDNA library from lactating 
guinea-pig mammary gland poly(A)-containing RNA (10).

Construction of recombinant DNA plasmids and transformation of 
E.coli cells

dT-tailed cDNA was annealed with an equimolar amount of the 
plasmid pAT153 (11) which had previously been linearised by 
restriction at the EcoRI site and tailed with dA residues, as 
described by Craig et al. (10). The amount of cDNA used for 
transformations varied from 10-25 ng. The resulting chimeric 
plasmids were used to transform E.coli HB101 recA^- cells using 
category II physical containment conditions in accordance with 
the guidelines laid down by the British Genetic Manipulation 
Advisory Group. Transformants were then selected on agar plates 
containing ampicillin (10). Resistant colonies were subsequently 
picked onto fresh plates containing 100 µg/ml ampicillin and 
12.5 µg/ml tetracycline.

In situ colony filter hybridisation

Recombinants containing sequences complementary to abundant 
poly(A)-containing RNA from the lactating human mammary gland 
were selected by in situ colony filter hybridisation, using base 
cleaved poly(A)-containing RNA isolated from either human or 
guinea-pig lactating tissue, and 32P-labelled to high specific 
radioactivity (about 10^7 cpm/µg) with T4 polynucleotide kinase 
(12). Hybridisation conditions were as described previously (10). 

Positive hybridisation translation

Recombinant plasmid DNA (20 µg) was partially digested with 
the restriction endonuclease HaeIII and then bound to 1 cm 
diameter discs of freshly prepared diazobenzyloxymethyl (DBM)- 
paper (13,14). Positive hybridisation translation was then 
carried out as described previously (10) except that both high 
stringency (20 mM NaCl, 8 mM sodium citrate pH 7.6, 0.2% (w/v) 
SDS, 60% (v/v) formamide) and low stringency (0.6 M NaCl, 20 mM, 
Hepes, pH 7.6, 1 mM EDTA, 0.2% (w/v) SDS, 60% (v/v) formamide) 
filter washes were used depending on the source of RNA. Eluted 
RNA was translated in a wheat germ cell-free protein-
synthesizing system in the presence of 35S-methionine (15), and
the products separated by SDS/polyacrylamide gel electrophoresis (16) and visualised by fluorography (17).

**Purification of a-lactalbumin cDNA using recombinant plasmids containing the a-lactalbumin sequences**

a-Lactalbumin recombinant plasmid DNA was partially restricted with HaeIII, heat denatured then immobilised on DBM-paper (10,13). This was then incubated for 16 h at 25°C with $^3$H-labelled cDNA (1.45 x 10$^7$ cpm/µg) complementary to poly(A)-containing RNA derived from either human or guinea-pig lactating mammary gland (7,15) in a buffer containing 0.6 M NaCl, 1 mM EDTA, 20 mM Hepes, pH 7.6, 0.1% (w/v) SDS, 20 µg/ml poly A, 60% (v/v) formamide. Filters were then washed individually with 4 x 1 ml aliquots of 20 mM NaCl, 8 mM sodium citrate, pH 7.6, 1 mM EDTA, 0.2% (w/v) SDS, 60% (v/v) formamide at 25°C for 5 min each. Hybridised cDNA was finally eluted with 3 x 100 µl aliquots of 0.1 M NaOH, 1 mM EDTA, at 37°C for 5 min each. After neutralisation with 1 M sodium acetate, pH 4.5, the cDNA was concentrated by ethanol precipitation.

**RNA·cDNA hybridisation**

Hybridisations were carried out at 68°C in 0.6 M NaCl, 50 mM Hepes, pH 7.4, 1 mM EDTA, 0.2% (w/v) SDS, in individual sealed siliconised glass capillary tubes. Following hybridisation for a suitable time (10s to 96h) the extent of hybridisation was determined using S$_1$ nuclease (15). For complex RNA populations best line fit was by computer as described previously (19).

**RESULTS**

**Construction of recombinant cDNA plasmids using total poly(A)-containing RNA isolated from lactating human mammary tissue**

All manipulations followed the procedure outlined in detail elsewhere (10) for the construction of recombinant plasmids containing guinea-pig milk protein cDNA sequences using our own modifications of many of the methods described by Kay et al., (12). Briefly, cDNA representative of the total poly(A)-containing RNA population of lactating human mammary tissue was synthesized using avian-myeloblastosis-virus reverse transcriptase. Size analysis using agarose gel electrophoresis (10) showed that the majority of these molecules ranged from 400-
2000 nucleotides in length. These were converted to a double stranded form using AMV reverse transcriptase, followed by S1 nuclease excision of the resultant hairpin loop. Analysis of this population after denaturation showed that the majority had a size distribution similar to that of the single stranded cDNA population, but that significant amounts of lower molecular weight material were also present. The latter were removed by preparative agarose gel electrophoresis (10), and a final double stranded cDNA population ranging in size from 350-2500 nucleotides was tailed with poly(dT)140Av using calf thymus terminal transferase. This was then annealed with EcoRI restricted pAT153 DNA which had been previously tailed with poly(dA)135Av. The resulting chimeric molecules were used to transform a recA- strain of E. coli HB101 using standard techniques (10) and the transformants selected on agar plates containing ampicillin. A transformation frequency of 2.8 x 10^3 colonies/μg of chimeric plasmid was obtained (equivalent to 2.3 x 10^4 colonies/μg of double-stranded cDNA), compared with 0.8 x 10^3 transformants/μg for tailed plasmid alone, and 3.6 x 10^5 transformants/μg for unrestricted pAT153. In total 767 transformants were obtained starting from 10 μg of poly(A)-containing RNA.

Selection of recombinant plasmids containing cDNA sequences that correspond to abundant poly(A)-containing RNA isolated from human lactating mammary tissue

Analysis of the kinetics of hybridisation of cDNA synthesized from total poly(A)-containing RNA derived from human lactating mammary gland showed that hybridisation occurred over six logarithmic decades, and that the best line fit by computer resolved this population into four abundance groups (see 18).

The very abundant fraction represented 10% of the total poly(A)-containing RNA (Fig. 1), and when compared with a highly purified rabbit globin mRNA kinetic standard (15), had a complexity of 1940 nucleotides and therefore was representative of 1 or 2 RNA species. On the basis of the source of tissue and the previous identification of α-lactalbumin and casein mRNA within the poly(A)-containing RNA population as determined by cell-free protein synthesis (7), it seemed reasonable to suppose
Figure 1. Complexity analysis and abundance of α-lactalbumin sequences in poly(A)-containing RNA from lactating human mammary gland. RNA excess hybridisations were performed as described in Methods. Complexity of total poly(A)-containing RNA from lactating human mammary gland versus its total homologous cDNA (○), or human α-lactalbumin cDNA (●); hybridisation of highly purified rabbit globin mRNA to its homologous cDNA (△); and hybridisation of human placental poly(A)-containing RNA to human α-lactalbumin cDNA (▲). Arrows denote RotJ values for different abundance groups determined by computer analysis of the data.

that the abundant poly(A)-containing RNA population represented the milk protein mRNA sequences. Consequently we have used the total human 32P-labelled poly(A)-containing RNA in in situ hybridisation studies (20) to determine in a preliminary manner which colonies were most likely to contain milk protein mRNA sequences. In addition, in view of the extensive sequence homology (69% identical amino acids) between human and guinea-pig α-lactalbumin (21,22), guinea-pig lactating mammary gland 32P-labelled poly(A)-containing RNA was also used to probe a duplicate set of filters. Of the 767 colonies screened in this manner, 46 (6%) showed significant hybridisation over background, the majority of these being positive with both probes.

Identification of recombinant plasmids containing human α-lactalbumin sequences

Sixteen of the 46 positive colonies were selected, grown in liquid culture, and the plasmid DNA isolated and digested with
several restriction endonucleases. Comparative analysis of the digestion patterns with the parental plasmid, showed that all but one contained additional DNA (results not shown). All those containing additional DNA were partially restricted with HaeIII, denatured, and bound to DBM-paper filters. These were then hybridised to cytoplasmic poly(A)-containing RNA obtained from the lactating guinea-pig mammary gland. Non-hybridisable material was then removed with the low stringency wash buffer and specifically bound RNA was eluted and translated in a wheat germ cell-free protein synthesizing system. It can be seen from Fig. 2 that a number of these clones do indeed contain an \(\alpha\)-lactalbumin sequence; in all, nine out of the fifteen examined

![Figure 2.](image)

*Figure 2.* Identification of recombinant plasmids containing human \(\alpha\)-lactalbumin sequences. Positive hybridisation translation assays were performed as described in Methods. Track (a), No added RNA; (b), poly(A)-containing RNA from lactating guinea-pig mammary gland; (c) to (g), guinea-pig poly(A)-containing RNA hybridised to DBM-paper bound pAT153 (c), phB-35 (d), phQ-32 (e), phQ-29 (f), and phO-53 (g); (h), human poly(A)-containing RNA hybridised to DBM-paper bound phB-35; (i), poly(A)-containing RNA from lactating human mammary gland.
hybridised with $\alpha$-lactalbumin mRNA. In the case of two of these recombinants, phB-35 and phO-53, this has been confirmed using human lactating mammary gland poly(A)-containing RNA and a more stringent washing procedure (Fig. 2).

**Restriction analysis of recombinant plasmids containing human $\alpha$-lactalbumin sequences**

Detailed restriction analysis of the recombinant phO-53 revealed that the $\alpha$-lactalbumin sequence in this plasmid contained one or more cleavage sites for the restriction endonucleases EcoRI, HindIII, AluI, HaeIII, and SstI, but no sites for BamHI, PstI, HpaII, HincII, XhoI or HhaI. To facilitate mapping of the inserted DNA, the phO-53 plasmid was first digested with HpaII and the large fragment containing the $\alpha$-lactalbumin sequence purified by sucrose gradient centrifugation. Subsequent digestion of this fragment with the enzymes EcoRI, HindIII, AluI, HaeIII, and HhaI, either singly or in combination (Fig. 3), led to the restriction map shown in Fig. 4. Confirmation of the ordering of fragments was obtained by analysis of different recombinants containing different regions of the $\alpha$-lactalbumin sequence (Fig. 4). It was also apparent that phO-53, phQ-32, and phQ-29 contained the $\alpha$-lactalbumin sequence in one orientation with respect to the plasmid, whereas phB-35, phP-52, and phD-30 had the opposite orientation.

From Fig. 4 it can be seen that phO-53 and phB-35 contain inserted DNA sequences of about 860 bp and 850 bp respectively (including the homopolymeric tails). Treatment of these recombinants with $S_1$ nuclease under partially denaturing conditions (28) resulted in excision of the inserted sequence and digestion of the A-T tails. This provided a better estimate of the size of the inserted cDNA sequence (minus tails) and a value of 670 bp was obtained for phB-35 and 680 bp for phO-53 (results not shown), suggesting a tail length of about 90 residues on each end of the inserted cDNA sequence.

When attempts were made to obtain a detailed restriction map of phB-35 DNA it became apparent that additional minor bands were present suggestive of plasmid heterogeneity within this clone. We have investigated this in the following manner.
Figure 3. Restriction analysis of a recombinant plasmid containing a human α-lactalbumin sequence. The large HpaII fragment containing the entire inserted sequence of phO-53 was purified by sucrose gradient centrifugation and then digested by a number of restriction endonucleases, either singly or in combination, and electrophoresed on a 1.6% agarose gel. Track (a), large HpaII fragment; (b), digestion with EcoRI; (c), HindIII; (d), HhaI; (e), HaeIII; (f), AluI; (i), HhaI/HindIII; (j), HaeIII/EcoRI; (k), AluI/EcoRI; (l), HhaI/HindIII; (m), HhaI/EcoRI. Tracks (g) and (h) contain AluI restricted pBR-322 and HaeIII restricted pAT153 markers respectively.

E. coli HB101 recA^- cells were transformed with phB-35 DNA and a number of the resulting colonies examined by restriction analysis. From this it seemed probable that the observed heterogeneity in the original phB-35 clone had arisen by a deletion process. Furthermore this deletion was always in a terminal fragment when subjected to restriction analysis. For example, DNA from three different clones containing phB-35 DNA (designated phB-35a,c, and d) were digested with HhaI and the large fragment containing the whole inserted sequence flanked by 103 bp of plasmid DNA was isolated by sucrose gradient
Figure 4. Restriction maps of human α-lactalbumin cDNA inserted sequences. The upper figure shows a detailed map of the inserted sequence of phO-53, together with the assigned location and orientation of the coding region. The lower figure compares the inserted sequences of a number of different human α-lactalbumin recombinants, and shows their orientations relative to the plasmid's unique HindIII site.

After centrifugation. When these purified fragments were digested with AluI, the internal fragments were common to all plasmids, but one terminal fragment varied in the cases of phB-35a and phB-35c, and both terminal fragments were altered in the case of phB-35d (Fig. 5). In the latter situation some heterogeneity in the length of the inserted sequence was still apparent suggesting that a further deletion had occurred during or subsequent to the second transformation.
Figure 5. Demonstration of plasmid heterogeneity in the phB-35 clone. The large HhaI fragment containing the entire inserted sequence flanked by 103 bp of plasmid DNA was purified from the original phB-35 plasmid DNA (track b) and from a number of its re-transformants (track c, phB35a; track d, phB35c; track e, phB35d). These were then digested with AluI (tracks f to i respectively). Track (a) contains HaeIII restricted pAT153.

In view of the terminal location of the observed deletions the most likely explanation seemed to be "looping out" of the homopolymeric A-T tails. To test this possibility the large HhaI fragments containing the inserted sequence of plasmids phB-35a, phB-35c and phB-35d were treated with S1 nuclease under partially denaturing conditions (28) resulting in excision of the tails. Although the lengths of these fragments before S1 treatment were clearly different (Fig. 6), after treatment they were identical, indicating that the heterogeneity resided in the A-T tails.

**Quantitation of human α-lactalbumin poly(A)-containing RNA sequences in lactating tissue**

Total $^3$H-labelled cDNA (1.45 x 10$^7$cpm/μg) was prepared using
Figure 6. Heterogeneity in the A-T tail lengths of phB-35 plasmid DNAs. The large HhaI fragment containing the entire inserted sequence was isolated from phB-35a (track b), phB-35c (track c) and phB-35d (track d). 1µg of each of these was then incubated with 150 units of S1 nuclease at 50°C for 15 min in 10µl of 0.2M NaCl, 1.5mM ZnSO4, 25mM sodium acetate, pH 4.6, containing 45% (v/v) formamide (tracks e to g). Track (a) contains AluI restricted pBR322.

poly(A)-containing RNA isolated from lactating human mammary tissue as the template. Human α-lactalbumin specific cDNA was then purified from this by hybridisation to phB-35 plasmid DNA immobilised on DBM-paper filters (see Materials and Methods). The purity and specificity of this α-lactalbumin cDNA was then determined by hybridisation to a large excess of partially restricted, heat-denatured phB-35 plasmid DNA in solution, under identical conditions to those used for RNA complexity analysis. It can be seen from Fig. 7 that hybridisation followed the expected first order kinetics of a pure probe, with hybridisation of the cDNA going to completion. When the α-lactalbumin cDNA was challenged with recombinant DNA from a plasmid known not to contain an α-lactalbumin sequence.
Figure 7. Specificity of purified human α-lactalbumin cDNA. Human α-lactalbumin cDNA was prepared as described in the Methods and then hybridised to an excess of phB-35 DNA (●), phC-22 DNA (○), and pls-1 DNA (▲), under identical conditions to those used for RNA complexity analysis.

(e.g. phC-22) S1 protection did not exceed 7% over a similar range of C0t values (Fig. 7), an indication that neither the pAT153 sequence nor the A-T tails bound significant levels of cDNA. Similarly, no protection was observed using the characterised plasmid pls-1 which contains the chick lysozyme cDNA sequence (23). Having established the purity of the α-lactalbumin cDNA probe, this was then used to determine the abundance of α-lactalbumin RNA sequences in a poly(A)-containing RNA preparation isolated from 15 day post-partum lactating mammary tissue, by RNA excess hybridisation. It can be seen from Fig. 1 that α-lactalbumin cDNA hybridised to this RNA with the expected first order kinetics of a pure probe, with hybridisation approaching completion. Furthermore the R0t50 value was similar to that of the very abundant RNA fraction identified by complexity analysis of the total poly(A)-containing RNA population (Fig. 1).

As a control, purified α-lactalbumin cDNA was also
challenged with total poly(A)-containing RNA isolated from human placental tissue. In this case $S_1$ nuclease protection did not exceed 20%, even at high $R_0$ values (Fig. 1).

**Analysis of sequence divergence between α-lactalbumin mRNAs from different species**

Analysis of the primary structure of α-lactalbumin from guinea-pig (22) and human (21), and lysozyme from chicken (29) has shown that significant homology exists within these amino acid sequences.

Our studies (Fig. 2) show that nucleotide sequence homology exists between human and guinea-pig α-lactalbumin mRNA species, since guinea-pig α-lactalbumin mRNA will bind efficiently to recombinant plasmid DNA which contains the human α-lactalbumin cDNA sequence. To examine the extent of nucleotide sequence homology between α-lactalbumin mRNAs from various species, we have used rhesus monkey α-lactalbumin cDNA. This was chosen since the lactating rhesus mammary tissue was more readily obtainable than the equivalent human tissue, whilst the respective α-lactalbumin mRNA species showed greater than 90% sequence homology (Fig. 8).

Rhesus α-lactalbumin cDNA was hybridised in RNA excess to poly(A)-containing RNA isolated from human, rhesus monkey, cynomolgus monkey, guinea-pig, rat and rabbit lactating mammary tissues. The homologous hybridisation with rhesus RNA led to about 95% protection of the probe from $S_1$ nuclease digestion, by a $R_0$ value of 1 mol.nuc$^{-1}$.l$^{-1}$.s$^{-1}$. Similar protection was obtained using cynomolgus RNA, and slightly less for human RNA (Fig. 8), indicating extensive nucleotide sequence homology between primate α-lactalbumin mRNA sequences. In contrast, hybridisation to guinea-pig, rat and rabbit RNA gave only 45%, 40%, and 30% protection respectively against $S_1$ nuclease digestion, indicating significant sequence divergence between primate and rodent α-lactalbumin sequences, as compared to human placental poly(A)-containing RNA which gave only 15% protection under identical conditions. Additional studies using guinea-pig α-lactalbumin cDNA and poly(A)-containing RNA from lactating human, guinea-pig, rat and rabbit mammary tissue, confirmed that significant homology exists between human and guinea-pig.
Figure 8. Analysis of sequence divergence between α-lactalbumin mRNAs from different species. RNA excess hybridisations were performed between purified rhesus monkey α-lactalbumin cDNA (solid lines) or guinea-pig α-lactalbumin cDNA (broken lines) and total poly(A)-containing RNA isolated from lactating human (▲), rhesus monkey (●), cynomolgus monkey (○), guinea-pig (▲), rat (▼) or rabbit (▼) mammary gland, or human placenta (□). RNA preparations were prepared from frozen tissue samples as described previously (7).

α-lactalbumin mRNA sequences (up to 60% protection), but demonstrated that extensive sequence divergence had occurred between guinea-pig, and rat and rabbit (20% protection, see Fig. 8). This somewhat surprising observation, was not a reflection of a degraded RNA population, as all RNA preparations used were active in cell-free protein synthesizing systems, and directed the synthesis of the expected spectrum of milk proteins (results not shown).

It would appear that hybridisations are specific to α-lactalbumin sequences, since (a) all heterologous hybridisations followed first-order kinetics suggesting hybridisation to a single RNA species; and (b) the $R_o \times 10^3$ value obtained for the guinea-pig cDNA versus human RNA cross hybridisation (Fig. 8) was identical to that for the homologous human hybridisation (Fig. 1), a demonstration that the rate of hybridisation, even
in heterologous cross-hybridisation studies, reflects the relative concentration of α-lactalbumin mRNA within a given poly(A)-containing RNA population.

DISCUSSION

We have described the construction of a library of recombinant plasmids containing cDNA sequences representative of total poly(A)-containing RNA isolated from lactating human mammary tissue (7). From this we have identified a series of recombinants which contain α-lactalbumin cDNA sequences. No recombinants containing casein cDNA sequences were identified, in spite of the apparent abundance of casein mRNA in the initial poly(A)-containing RNA preparation as previously determined by cell-free protein synthesis (7), an observation in common with the unusually low frequency of recombinants containing guinea-pig casein C cDNA sequences (10). The latter we have surmised is a result of secondary structure within the casein C mRNA due to its predicted high G+C content. This secondary structure may well affect the efficiency of the reverse transcriptase (24).

It is possible that the low frequency (or absence) of human casein cDNA recombinants may also be a reflection of extensive secondary structure within the mRNA. The human major or β-casein (25) in common with guinea-pig casein C (26) contains an unusually high proportion of proline, glutamic acid, valine and leucine residues, whilst both caseins share with bovine β-casein (27) amino acid sequence homology at the NH2-terminal end (R. Craig - unpublished). Irrespective of codon usage, mRNA molecules encoding a preponderance of proline and glutamic acid residues will have regions of high G+C content, and consequently result in regions of stable secondary structure. Moreover, if the preference for C or G in the third eucaryote codon prevails (see 30) then regions encoding valine and leucine will also be G+C rich.

Restriction enzyme analysis of a series of human α-lactalbumin cDNA recombinants identified plasmids with the cDNA sequence inserted in either orientation. Two of these (phB-35 and phO-53) probably contain the complete coding sequence for the primary translation product (pre-α-lactalbumin
This was determined by alignment of the restriction map (Fig. 4) with the amino acid sequence (21). The restriction enzymes EcoRI, HindIII and AluI were particularly useful. The single EcoRI site must lie within the non-coding region of the mRNA since no cleavage site may be assigned from the known amino acid sequence. We have assumed on the basis of its position when related to the size of the inserted cDNA that this will lie within the 3' non-coding region of the mRNA. This assumption seems to be correct, since it permits alignment of the AluI and HindIII sites. These can cleave at sequences encoding Lys-Leu. Since this amino acid combination occurs only at residues 58/59 and 114/115 of the polypeptide, 168 bases apart, it seems likely that they represent the fragment we estimate to be 165 base pairs generated by an AluI digest. This we can confirm by alignment of the adjacent AluI fragment which we estimated to be of 140 base pairs. This fragment is also predictable from the amino acid sequence, and represents cleavage at the Gln-Leu (residues 10/11) by AluI and at the Lys-Leu (residues 58/59) by either AluI or HindIII, and should account for 144 bases.

Protein sequencing studies have shown that the mutation rate among α-lactalbumins is high, and that rabbit α-lactalbumin varies more from other sequenced α-lactalbumins than the latter vary from each other (31). Comparison of the amino-acid sequence of rabbit (31) and guinea-pig (22) reveals 58 amino-acid substitutions, less than half the amino acid complement. However, the C-terminal sequence is well conserved, a feature common to all sequenced α-lactalbumins (see 31). Thus in the C-terminal sequence of 29 amino acids in rabbit α-lactalbumin, 20 are conserved in guinea-pig α-lactalbumin, and of the remainder, 7 differences may be accounted for by a single base change. If this amino-acid sequence homology was reflected within the nucleotide sequence due to identical codon usage, then we would expect some nucleotide sequence homology between guinea-pig and rabbit α-lactalbumins. This however is not the situation. Moreover, sequence divergence must extend into the 3' untranslated region of the two mRNA species, since cDNA probes will inevitably be enriched in 3'-terminal sequences.
Thus in rodents, as opposed to primates, hybridisation studies suggest that divergence of α-lactalbumin genes is greater than previously envisaged from protein sequencing studies alone.

Cross-hybridisation studies provide an additional interesting feature of nucleotide sequence divergence within the α-lactalbumin gene family. Although there is marked sequence divergence between rodent sequences, all show significant sequence homology with primate α-lactalbumin sequences. Thus, whereas little sequence homology is apparent between guinea-pig and rat α-lactalbumin mRNA, both show 40-45% sequence homology with rhesus monkey mRNA, an observation compatible with the concept that the rat and guinea-pig sequences show nucleotide sequence homology with different regions of the rhesus monkey α-lactalbumin mRNA.

Guinea-pig (10), human and rat (32) α-lactalbumin cDNA sequences have now been cloned. Comparison of the nucleotide sequence of these different α-lactalbumin mRNA species should provide some insight into the sequence divergence which has occurred within the α-lactalbumin gene during evolution.

ACKNOWLEDGEMENTS

We thank the Cancer Research Campaign for supporting this work, Professor P.N. Campbell, Dr. S. Jupp and Dr. J.G. Williams for many valuable discussions, and Miss J. Allison for technical assistance. pAT153 and pls-1 DNA were the gifts of Professor D. Sherratt and Dr. A.E. Sippel. We are also grateful to Mr. P. Squires of the Huntingdon Research Station for cynomolgus monkey mammary tissue, and Dr. J. Marston, Department of Anatomy, University of Birmingham for the rhesus monkey mammary tissue. AMV reverse transcriptase was the gift of Dr. J. Beard and Dr. J. Gruber.

REFERENCES