Biochem. J. (1981) **194**, 989–998
Printed in Great Britain

The construction, identification and partial characterization of plasmids containing guinea-pig milk protein complementary DNA sequences

Roger K. CRAIG,* Len HALL, David PARKER and Peter N. CAMPBELL
Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, Mortimer Street, London W1P 7PN, U.K.

(*Received 22 October 1980/Accepted 19 December 1980*)

A complementary DNA (cDNA) plasmid library has been constructed in the plasmid pAT153, using poly(A)-containing RNA isolated from the lactating guinea-pig mammary gland as the starting material. Double stranded cDNA was inserted into the *Eco*RI site of the plasmid using poly(dA·dT) tails, then transformed into *Escherichia coli* HB101. From the resulting colonies we have selected and partially characterized plasmids containing cDNA copies of the mRNAs for casein A, casein B, casein C and α-lactalbumin. However, the proportion containing casein C cDNA was exceptionally low, and these contained at best 60% of the mRNA sequence.

The expression of the milk protein genes by the mammary gland is known to be modulated by an intricate combination of steroid and peptide hormones (see Banerjee, 1976). Studies using several animal systems have shown that the regulatory mechanisms are complex, and involve transcriptional, post-transcriptional and translational control (Guyette *et al.*, 1979; Bathurst *et al.*, 1980*a*,*b*; Craig *et al.*, 1980; Burditt *et al.*, 1981). Other studies have demonstrated that, in common with most proteins destined to be secreted, the milk protein mRNAs encode precursor polypeptides (see Craig *et al.*, 1978*b*). However, studies on the organization and structure of the milk protein genes have yet to be performed, due to the absence of suitable DNA hybridization probes.

Here we describe (i) the construction of a library of recombinant cDNA plasmids, containing sequences complementary to poly(A)-containing RNA isolated from the lactating guinea-pig mammary gland, and (ii) the selection and partial characterization of plasmids from this library containing wholly or in part the coding sequences for guinea-pig caseins A, B, and C, and the whey protein, α-lactalbumin (see Craig *et al.*, 1976, 1978*a*, 1979).

Abbreviations used: cDNA, complementary DNA; SDS, sodium dodecyl sulphate; DBM-, diazobenzyloxymethyl-; Hepes, 4-(2-hydroxyethyl)-1-piperazine-ethanesulphonic acid.

* To whom correspondence should be addressed.

Materials and methods

Materials

Escherichia coli DNA polymerase I and restriction enzymes *Alu*I, *Hae*III, *Eco*RI and nuclease-free bovine serum albumin were obtained from Bethesda Research Laboratories Inc. through Uniscience, Cambridge CB5 8BA, U.K.; T4 polynucleotide kinase was from P.L. Biochemicals and S_1 nuclease and deoxyribonuclease I were from Sigma. L-[^{35}S]-Methionine (700–1300 Ci/mmol), deoxy[5-^{3}H]-cytidine 5′-triphosphate (22 Ci/mmol), [*methyl*-^{3}H]-thymidine 5′-triphosphate (47 Ci/mmol), deoxy[8-^{3}H]adenosine 5′-triphosphate (14.6 Ci/mmol) and adenosine 5′-[γ-^{32}P]triphosphate (2000–3000 Ci/mmol) were from The Radiochemical Centre; avian-myeloblastosis-virus reverse transcriptase (lot no. G-379) was provided by Dr. J. W. Beard, Life Sciences Inc., St. Petersburg, FL 33707, U.S.A.; calf thymus terminal deoxynucleotidyl transferase was kindly provided by Dr. I. R. Johnston (University College, London, U.K.), restriction nuclease *Hpa*II by Dr. B. Griffin (I.C.R.F. Laboratories, London, U.K.), pAT153 DNA by Professor D. Sherratt (University of Glasgow, Glasgow, U.K.), and pBR322 DNA by Dr. J. F. Williams (I.C.R.F. Laboratories, Mill Hill, London, U.K.); all other chemicals and solvents were obtained from sources previously described (Craig *et al.*, 1976, 1979; Bathurst *et al.*, 1980*a*,*b*). Materials for bacterial growth were obtained from Difco.

0306-3283/81/030989-10$01.50/1

Preparation of solutions and glassware

Preparation of buffers, scintillation fluids, and glassware was as described previously (Craig *et al.*, 1979), except that diethylpyrocarbonate was only included in buffers required for RNA isolation procedures.

Buffers and medium

Buffer H was 10mM-Tris/HCl, pH7.4 at 20°C, containing 100mM-NaCl, 1mM-EDTA and 0.1% SDS.

SSC was 15mM-sodium citrate (pH7.0) containing 150mM-NaCl. L-Broth contained 1% (w/v) tryptone, 0.5% yeast extract, 0.5% NaCl and 4mM-glucose, pH7.2.

Isolation of poly(A)-containing RNA from lactating guinea-pig mammary gland

Total post-nuclear poly(A)-containing RNA was isolated from the lactating guinea-pig mammary gland essentially as described previously (Craig *et al.*, 1979), except that the total RNA was incubated with 10μg of iodoacetylated deoxyribonuclease/ml (Zimmerman & Sandeen, 1966) for 10min on ice in 10mM-Tris/HCl, pH7.4, containing 10mM-NaCl and 10mM-magnesium acetate, extracted once with phenol/chloroform (1:1, v/v) and the RNA was concentrated by ethanol precipitation before affinity chromatography on oligo(dT)–cellulose (Craig *et al.*, 1976).

Synthesis of poly(dT)-tailed double-stranded cDNA

Low specific radioactivity ^{3}H-labelled DNA (720c.p.m./ng) complementary to total post-nuclear poly(A)-containing RNA was synthesized as described previously (Craig *et al.*, 1979) with the modification that all four deoxynucleoside triphosphates were at a concentration of 500μM and [5-^{3}H]dCTP at 0.5mCi/ml. Single-stranded cDNA was then dissolved at a concentration of 1μg/ml (2ml assay) in 50mM-Tris/HCl, pH8.3 at 37°C, containing 10mM-dithiothreitol and 9mM-$MgCl_2$, with all four deoxyribonucleoside triphosphates each present at 500μM (see Kay *et al.*, 1980). The second strand was then synthesized by incubation at 37°C for 1h in the presence of 100U of avian-myeloblastosis-virus reverse transcriptase/ml. The reaction was stopped by the addition of EDTA and SDS to concentrations of 20mM and 1% (w/v) respectively. *E. coli* tRNA (10μg) was then added, the mixture extracted with phenol/chloroform (1:1, v/v) and the nucleic acid was precipitated with ethanol and recovered by centrifugation as described previously (Craig *et al.*, 1979). It was then dissolved in a small volume of double distilled water and residual deoxynucleoside triphosphates and traces of phenol were removed by gel filtration on a Sephadex G-100 column (0.8cm × 15cm) pre-equilibrated and eluted with 50mM-sodium acetate, pH4.5, containing 0.3M-NaCl and 3mM-$ZnSO_4$. Material eluted in the void volume was pooled, and single-stranded material was degraded by digestion with S_1 nuclease (2000U/ml) at 37°C for 20min. The reaction was stopped by the addition of EDTA and SDS to concentrations of 20mM and 1% (w/v) respectively, the reaction mixture was extracted once with phenol/chloroform (1:1, v/v) and the resulting aqueous phase was fractionated on a Sephadex G-100 column pre-equilibrated and eluted in buffer H. Nucleic acid eluted in the void volume was concentrated by ethanol precipitation in the presence of 20μg of *E. coli* tRNA.

Double-stranded cDNA was then fractionated on the basis of size by electrophoresis on a 1.2% (w/v) agarose horizontal slab gel (McDonell *et al.*, 1977) run in 0.2M-glycine, adjusted to pH8.5 with NaOH, at 200V (approx. 40mA) for 90min. The double-stranded cDNA was loaded in a 1cm central slot, and outside slots were loaded with restricted plasmid DNA for use as molecular weight markers. When electrophoresis was complete, regions of the gel containing the DNA marker fragments were sliced off and soaked in ethidium bromide (1μg/ml). The fragments were visualized with a u.v. transilluminator and their relative position was used to determine the region of the gel containing cDNA fragments ranging in size from 400 to 1650 base pairs. This region was excised, finely chopped, passed through a 21-gauge syringe needle into an equal volume of 10mM-Tris/HCl, pH8.0, containing 0.15M-NaCl and 1mM-EDTA (Boseley *et al.*, 1978), and allowed to stand overnight at 4°C. Agarose gel fragments were then sedimented by centrifugation at 200000$g_{av.}$ for 30min at 4°C in the SW56 rotor of a Beckman L2-65B ultracentrifuge. cDNA in the supernatant was concentrated by ethanol precipitation in the presence of 20μg of *E. coli* tRNA, dissolved in double distilled water and stored at −70°C. Recovery represented 60% of the initial cDNA applied to the gel.

Size-selected double-stranded cDNA was tailed with dT residues by using terminal deoxynucleotidyl transferase in the following manner. Double-stranded cDNA (400ng/ml) was dissolved in 30mM-Tris/100mM-cacodylic acid (adjusted to pH7.5 with 10M-KOH) containing 1mM-$CoCl_2$, 0.1mM-dithiothreitol, 50μg of nuclease-free bovine serum albumin/ml, and 0.1mM-[^{3}H]dTTP (2Ci/mmol). Terminal deoxynucleotidyl transferase (75 units/ml; one unit incorporates 1nmol of dATP per hour at 37°C) was then added and the mixture was incubated at 30°C. When a 100–150 residue long poly(dT) tail had been added (approx. 15min as determined by trichloroacetic acid precipitation of 1μl amounts at 5min intervals), the reaction was stopped by extraction with phenol/chloroform and

the nucleic acid was recovered by ethanol precipitation as described above. The precipitate was washed twice with ethanol, dissolved in double distilled water and stored at −70°C.

Preparation of poly(dA)-tailed, EcoRI-restricted pAT153

Supercoiled pAT153 DNA (300μg) was digested with 600 units of *Eco*RI at 37°C for 150min in 800μl of 100mM-Tris/HCl, pH7.5, containing 50mM-NaCl, 10mM-$MgCl_2$ and 100μg of nuclease-free bovine serum albumin/ml. The reaction was stopped by extraction with phenol/chloroform, and the plasmid DNA in the aqueous phase was precipitated with ethanol. Residual low molecular weight DNA was then removed by further purification by using sucrose gradient centrifugation in the following manner. The DNA was dissolved in 200μl of 10mM-Tris/HCl, pH7.6 at 4°C, containing 200mM-NaCl and 10mM-EDTA. This was applied to the top of a 12ml 5–20% (w/v) linear sucrose gradient containing the same buffer and centrifuged at 160000$g_{av.}$ for 16h at 4°C in the SW40 rotor of a Beckman L5-65 ultracentrifuge. Peak fractions of A_{258} were pooled, concentrated by ethanol precipitation and used as a source of linear *Eco*RI-restricted pAT153 DNA, 40μg of which was then tailed with poly(dA) essentially as described above, except that the DNA was present at 80μg/ml, and [^{3}H]dATP (1Ci/mmol) replaced dTTP as the source of radiolabelled deoxynucleoside triphosphate.

Recombinant DNA plasmid construction, transformation and selection

Poly(dT)-tailed cDNA and poly(dA)-tailed pAT153 DNA were annealed in 10mM-Tris/HCl, pH7.6, containing 200mM-NaCl and 1mM-EDTA, in polypropylene 'snap-cap' tubes at a final concentration of 0.4μg/ml and 4μg/ml respectively. The mixture was heated in a water-bath to 60°C for 30min, the water-bath was switched off, and the hybridization mixture was allowed to cool to room temperature overnight. The amount of poly(dT)-tailed cDNA used for transformations varied from 10ng to 50ng.

The resulting chimeric plasmids were then used to transform *E. coli* HB101 rec A$^-$ (Boyer & Roulland-Dussoix, 1969) under category II* physical containment in the following manner.

E. coli HB101 rec A$^-$ was grown in L-broth to an A_{550} of 0.8. The cells were then sedimented by centrifugation at 3000$g_{av.}$ for 5min at 4°C, then resuspended in their original volume in ice-cold 5mM-Tris/HCl, pH7.6, containing 100mM-$CaCl_2$ and 5mM-$MgCl_2$, and allowed to stand on ice for 30min. The cells were resedimented, then resuspended in 0.1 vol. of $CaCl_2$-containing buffer, and the now competent cells were kept on ice until required (up to 24h).

Chimeric plasmid DNA [derived from 25ng of poly(dT)-tailed cDNA] was then diluted to a final volume of 500μl with 10mM-Tris/HCl, pH7.6, containing 1mM-EDTA and added to 250μl of the competent cell suspension. These cells were incubated on ice for 20min, and then at 37°C for 5min without shaking. L-broth (2.5ml) was then added, and the cell suspension was gently shaken for 60min at 37°C. Aliquots (100μl) were plated onto 1% (w/v) agar in L-broth containing 100μg of ampicillin/ml and incubated overnight at 37°C. Resistant colonies were then picked onto fresh 9cm agar plates and also replica picked onto 9cm Millipore filters overlaying agar. In this and all subsequent growth procedures, ampicillin and tetracycline were present at 100μg/ml and 12.5μg/ml respectively.

Colony filter hybridization in situ

The procedure represents a simplified version of that outlined by Grunstein & Hogness (1975) and has evolved from several laboratories. Colonies picked onto Millipore filters, as described above, were allowed to grow overnight at 37°C. The filters were then removed and placed at room temperature for 7min on several thicknesses of Whatman 3MM paper saturated wtih 0.5M-NaOH. The filters were then transferred to 3MM paper saturated with 1M-Tris/HCl, pH7.5, for 2min. This step was repeated, and finally the filters were placed for 5min on 3MM paper saturated with 0.5M-Tris/HCl, pH7.5, containing 1.5M-NaCl. The filters were transferred onto several thicknesses of dry 3MM paper, then each individually transferred to the suction apparatus described by Grunstein & Hogness (1975). A vacuum was applied to fix the colonies to the filter, and the colonies were washed under vacuum with 60ml of ethanol. All the filters were then air-dried overnight, and finally baked at 80°C for 2h. Up to four filters were placed upside down in a 9cm plastic petri dish containing 10ml of 5 × SSC, 50% (v/v) formamide, 0.1% SDS, 25μg of *E. coli* tRNA/ml and 10^6c.p.m./filter of ^{32}P-labelled base cleaved mRNA [sp. radioactivity (1–2) × 10^7c.p.m./μg] as the hybridization probe. The filters were incubated overnight at 37°C. All filters were then transferred to a beaker and the ^{32}P-labelled mRNA which had not hybridized was removed by successive washes with the following buffers. Filters were washed twice for 30min each at 37°C in 200ml of 5 × SSC containing 50% (v/v) formamide, once at room temperature for 30min in 100ml of 2 × SSC containing 1 unit of ribonuclease A/ml (boiled for 7min before use), and finally twice for 30min in 200ml of 2 × SSC. The filters were then air-dried, and autoradiographed at −70°C.

The ^{32}P-labelled mRNA probe was prepared by

base-cleavage of poly(A)-containing RNA (50 μg/ml) in 100 mM-NaOH at 4°C for 60 min. This was then neutralized by the addition of an appropriate volume of HCl followed by 0.1 vol. of 1 M-Tris/HCl, pH 8.0, and the RNA was recovered by ethanol precipitation. The RNA was then ^{32}P-end-labelled with T_4 polynucleotide kinase. Routinely, 2 μg of RNA was incubated with 300 μCi of [γ-^{32}P]ATP in a final volume of 45 μl. Each assay also contained 70 mM-Tris/HCl (pH 7.6)/10 mM-$MgCl_2$/5 mM-dithiothreitol, 100 μg of bovine serum albumin/ml, and 100 units of T_4 polynucleotide kinase/ml. The final assay mixture was incubated for 60 min at 37°C, and assay was terminated by extraction with an equal volume of phenol/chloroform. The supernatant phase was diluted to 500 μl with buffer H, then loaded onto a Sephadex G-50 column (0.8 cm × 16.0 cm) equilibrated and eluted with buffer H. ^{32}P-labelled RNA eluted in the void volume was then concentrated by ethanol precipitation in the presence of 20 μg of *E. coli* tRNA. The final precipitate was dissolved in a small volume of double distilled water and used immediately for hybridization *in situ*.

Amplification and isolation of plasmid DNA

Transformed bacteria were grown at 37°C with vigorous shaking in L-broth containing ampicillin (100 μg/ml) and tetracycline (12.5 μg/ml) to an A_{550} of 0.6–0.9. Chloramphenicol (170 μg/ml) was then added and incubation was continued for a further 18–20 h. The cells were recovered by centrifugation, cell lysates prepared, and plasmid DNA separated from chromosomal DNA by isopycnic banding in CsCl-density gradients containing ethidium bromide, as described by Katz *et al.* (1973). Ethidium bromide was removed from plasmid DNA by repeated extractions with isoamyl alcohol saturated with CsCl, the DNA was dialysed against 10 mM-Tris/HCl, pH 7.5, containing 1 mM-EDTA, then concentrated by ethanol precipitation.

All manipulations involving live cells were performed under conditions of category II* physical containment as recommended by the British Genetic Manipulation Advisory Group.

Size analysis of DNA preparations

^{3}H-labelled cDNA or plasmid DNA samples were heat denatured in sealed capillaries at 100°C for 10 min in 0.5% (w/v) SDS/20 mM-EDTA/0.5 mg/ml Bromophenol Blue, then electrophoresed on 1.2% (w/v) agarose horizontal slab gels run in 90 mM-boric acid/90 mM-Tris/1.25 mM-EDTA, pH 8.35, at 200 V (about 45 mA) for 2–3 h. After electrophoresis, parallel tracks containing heat-denatured, restricted plasmid DNA markers were stained for 1 h in ethidium bromide (1 μg/ml), and the band mobilities measured under u.v. illumination. Tracks containing ^{3}H-labelled DNA were cut into 2.5 mm slices, incubated at 50°C for 2 h with 0.75 ml of NCS Tissue Solubilizer, and counted after the addition of 10 ml of toluene scintillant.

Positive hybridization–translation

Recombinant plasmid DNA (15–20 μg) was partially restricted with *Hpa*II or *Hae*III and then covalently bound to 1 cm diameter discs of freshly prepared DBM-paper as described by Wahl *et al.* (1979). Up to four filters were incubated at 25°C for 16 h in 100 μl of 20 mM-Hepes, pH 7.6, containing 0.6 M-NaCl, 1 mM-EDTA, 0.2% SDS, 50 μg of wheat germ tRNA/ml, 20 μg of poly(A)/ml, 60% (v/v) formamide, and 4 μg of total post-nuclear poly(A)-containing RNA isolated from lactating guinea-pig mammary gland. Individual filters were then washed four times at 25°C for 10 min each with portions (3 ml) of 20 mM-NaCl/8 mM-sodium citrate (pH 7.6)/0.2% SDS/60% (v/v) formamide. Hybridized RNA was eluted at 37°C for 10 min with 100 μl of 20 mM-Hepes, pH 7.6, containing 1 mM-EDTA, 0.05% SDS, 100 μg of wheat germ tRNA/ml and 90% (v/v) formamide. The filters were then washed with two 100 μl portions of distilled water and the washes were added to the eluted RNA. The RNA was then recovered by ethanol precipitation and the mRNA species were identified by cell-free protein synthesis.

Size estimation of milk protein mRNA sequences by using the 'Northern' transfer technique

RNA samples were denatured by treatment with glyoxal, then separated on the basis of size by electrophoresis on 1.5% (w/v) agarose gels. Longitudinal strips of gel containing marker RNA species were removed and then visualized after soaking in Acridine Orange, all as described by McMaster & Carmichael (1977). The milk protein RNA was then transferred from the gel and covalently coupled to DBM-paper as described by Wahl *et al.* (1979), and the positions of the milk protein mRNA sequences were determined by using nick-translated recombinant plasmid DNA hybridization probes (Rigby *et al.*, 1977). Hybridization conditions and subsequent washing procedures were also as described by Wahl *et al.* (1979). The position of mRNA sequences was then visualized by autoradiography at −70°C.

Cell-free protein synthesis and product analysis

mRNA-directed cell-free protein synthesis was carried out in a nuclease-treated reticulocyte lysate as described by Pelham & Jackson (1976). SDS/polyacrylamide-gel electrophoretic analysis was performed by using slab gels as described previously (Craig *et al.*, 1979). Fluorography followed the procedure of Bonner & Laskey (1974) except that

dimethyl sulphoxide was replaced by glacial acetic acid throughout (Burckhardt *et al.*, 1979).

Restriction enzyme analysis

Unless specifically stated otherwise all restriction enzyme analyses were performed using the ionic conditions recommended by the manufacturers.

Results

Construction of recombinant derivatives of the plasmid pAT153 by using poly(A)-containing RNA isolated from the post-nuclear supernatant of the lactating guinea-pig mammary gland

cDNA representative of the total post-nuclear poly(A)-containing RNA population was synthesized by using avian-myeloblastosis-virus reverse transcriptase. Analysis of the size distribution of the final product by agarose-gel electrophoresis when compared with restriction fragments of known size demonstrated (Fig. 1*a*) that the majority of the cDNA molecules were between 400 and 1100 nucleotides in length. These sequences were then converted into a 'hairpin'-double-stranded DNA population by using avian-myeloblastosis-virus reverse transcriptase, and the extent of the reaction was monitored by S_1 nuclease digestion of a small amount of the final product. Routinely 45–55% of the input [^{3}H]cDNA was found to be resistant to S_1 nuclease. Conversion of the 'hairpin'-double-stranded DNA products into a truly double-stranded form was achieved by degradation of the 'hairpin' loop and the remaining single-stranded cDNA with high levels of S_1 nuclease (McReynolds *et al.*, 1977). This resulted in a population of sequences similar in size to the single-stranded population, though with a slight bias towards the smaller (400–600 base pair) sequences (Fig. 1*b*). To remove low molecular weight sequences, the total double-stranded DNA population was applied to an agarose gel, and the DNA which electrophoresed to a position corresponding to sequences of between 400 and 1650 base pairs was eluted, and recovered by ethanol precipitation. A small amount of this was then re-analysed by agarose-gel electrophoresis (Fig. 1*b*). This demonstrates that the preparation now consisted predominantly of sequences ranging in size from 400 to 1100 base pairs.

The size-selected double-stranded cDNA was then tailed with poly(dT) with calf thymus terminal transferase, and the reaction was stopped after the addition of an estimated 135 residues (this calculation assumes a mean size of 600 base pairs for the starting material). A small amount of the product was then analysed by agarose-gel electrophoresis. This showed (Fig. 1*c*) that the majority of the sequences were of relatively higher molecular weight (500–1100 base pairs) though a population of appar-

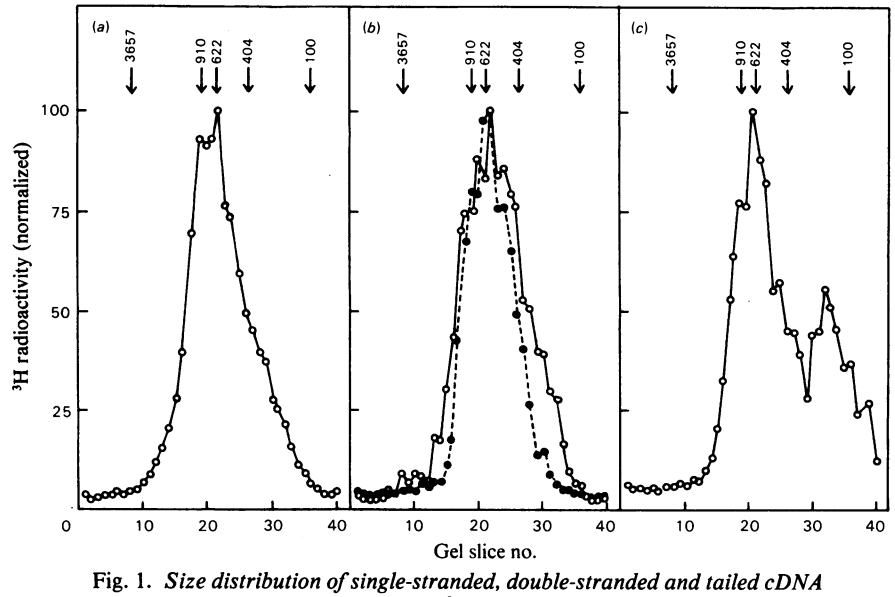

Fig. 1. *Size distribution of single-stranded, double-stranded and tailed cDNA*

In each instance about 5000 c.p.m. of heat-denatured ^{3}H-labelled cDNA was electrophoresed on a 1.2% (w/v) agarose gel, then sliced and assayed for radioactivity as described in the Materials and methods section. *Eco*RI- and *Hpa*II-restricted pAT153 and *Alu*I-restricted pBR322 were heat-denaturated and run in parallel tracks to provide suitable size markers (Sutcliffe, 1978*a*,*b*). (*a*) Single-stranded cDNA, (*b*) double-stranded cDNA before (○) and after (●) size selection on a preparative agarose gel, (*c*) tailed double-stranded cDNA.

ent lower molecular weight was also present (80–300 base pairs). Subsequent analysis demonstrated that this consisted almost entirely of residual [^{3}H]dTTP, and contained little acid-precipitable material (L. Hall & R. Craig, unpublished work). The final 'tailed' double-stranded cDNA population represented a weight yield of 8% when related to the amount of single-stranded cDNA used as starting material.

The $(dT)_{\overline{135}}$-tailed double stranded cDNA was then annealed overnight (see methods) with the plasmid pAT153 (Twigg & Sherratt, 1980), which had been previously restricted at the *Eco*RI site and then tailed with terminal transferase in the presence of [^{3}H]dATP until an estimated 136 residues had been added. The resulting chimeric molecules were used to transform an *E. coli* rec A$^-$ strain. Transformants were then selected on agar plates containing ampicillin. As controls, the transformation of unrestricted and $(dA)_{\overline{136}}$-tailed pAT153 were performed in parallel. This resulted in a transformation frequency of 3.6×10^6 colonies/μg for unrestricted pAT153, 0.8×10^3 colonies/μg for tailed pAT153, and 2.2×10^3 colonies/μg for recombinant DNA plasmids (or 2.2×10^4/μg of double stranded cDNA).

The recombinant colonies were replica-plated onto fresh plates and onto Millipore filters, and those containing the abundant milk protein sequences were identified by hybridization *in situ* with base-cleaved ^{32}P-labelled post-nuclear poly(A)-containing RNA as the probe. Of the 497 colonies plated, 131 hybridized to the ^{32}P-labelled RNA to a significant degree above the background. Thirty of these gave a particularly intense image on the autoradiograph, suggesting that these were the most likely to contain the abundant milk protein mRNA sequences.

Seventeen colonies which hybridized strongly with ^{32}P-labelled mRNA were grown in bulk (1 litre cultures) and the plasmid DNA was isolated. All but one contained additional DNA, by virtue of their decreased mobility when electrophoresed on agarose gels as compared with the parental plasmid pAT153 (results not shown). Although all have been extensively analysed, we describe in detail below the characterization of four plasmids, pgp N-33, pgp K-27, pgp O-22, and pgp K-9.

Identification and partial characterization of recombinant plasmids containing milk protein cDNA sequences

Comparative restriction analysis of the four plasmids with *Hpa*II (Fig. 2) demonstrates that all contained additional sequences to those present in the parental plasmid, and that two, pgp N-33 and pgp K-9, contain an additional *Hpa*II site within these sequences. Estimation of the size of the inserted DNA sequences from the *Hpa*II restriction

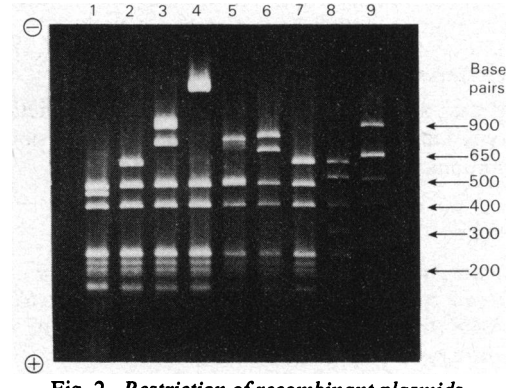

Fig. 2. *Restriction of recombinant plasmids*
Each recombinant plasmid (1.5 μg) was restricted with *Hpa*II, then electrophoresed on a 1.2% (w/v) agarose gel as described in the Materials and methods section. As a comparison pAT153 was restricted with *Hpa*II, and also with *Hpa*II and *Eco*RI. *Alu*I- and *Hpa*II-restricted pBR322 were electrophoresed in parallel tracks as a source of suitable size markers (Sutcliffe, 1978*a*,*b*). Lane 1, *Hpa*II/*Eco*RI digest of pAT153; lane 2, *Hpa*II digest of pAT153; lanes 3–6, *Hpa*II digests of pgp N-33, pgp K-27, pgp O-22 and pgp K-9 respectively; lane 7, *Hpa*II digest of pAT153; lane 8, *Hpa*II digest of pBR322; lane 9, *Alu*I digest of pBR322.

patterns suggested that the four plasmids, pgp N-33, pgp K-27, O-22, and pgp K-9, contained about 1300, 910, 660 and 820 base pairs of inserted sequences (including homopolymeric tails) respectively. One plasmid, pgp O-22, contained two additional minor bands, indicating the presence of other recombinant material. Subsequent detailed restriction analysis has shown that each contains casein C cDNA, and that the smaller inserted sequences represent deletions of the largest. We have observed a similar phenomenon in recombinant plasmids containing human milk protein sequences, constructed in an identical manner, and have shown that in this case the deletions are located solely in the poly(dA·dT) tail region (Hall *et al.*, 1981).

We have also assessed the size of the inserted sequences by S_1 nuclease excision, as described by Hofstetter *et al.* (1976). In this instance 1060, 850, 650 and 880 base pairs of inserted sequences were estimated to be present in the respective plasmids (results not shown). This probably represents a realistic estimate of the inserted cDNA sequence, since under these conditions the poly(dA·dT) tails are excised.

Identification of the coding sequences present within each plasmid was carried out by SDS/polyacrylamide-gel electrophoresis of the cell-free

translation products of mRNA isolated by hybridization to partially restricted denatured plasmid DNA immobilized on DBM-paper filters (see Smith *et al.*, 1979). In this instance, the purified mRNA sequences were translated in a reticulocyte lysate and the resulting [^{35}S]methionine-labelled polypeptides were analysed by using a phosphate-buffered SDS/polyacrylamide-gel system known to separate all the primary translation products of the three predominant guinea-pig caseins and α-lactalbumin (see Craig *et al.*, 1979). This demonstrated (Fig. 3) that the plasmids pgp N-33, pgp K-27, pgp O-22 and pgp K-9 contained wholly, or in part, the coding sequence for guinea-pig casein A, casein B, casein C, and α-lactalbumin respectively.

To determine the proportion of each mRNA within the characterized recombinants, the size of each mRNA was estimated by using RNA blotting techniques (Alwine *et al.*, 1977) after separation of poly(A)-containing RNA on the basis of size by electrophoresis under denaturing conditions (see the Materials and methods section). The results (Fig. 4) show that α-lactalbumin was the smallest mRNA and contained 800 ± 50 nucleotides, whereas all the casein mRNAs were 1050 ± 100 nucleotides in size. Thus, it seems that the cDNA inserted into pgp N-33 (casein A) and pgp K-9 (α-lactalbumin) represent large, if not full-length, copies of the respective mRNA sequences, pgp K-27 (casein B) up to 80% of the mRNA sequence, but pgp O-22 (casein C) only 55–60% of the mRNA sequence.

Identification of additional recombinant plasmids containing coding sequences for each of the four proteins was achieved in the following manner. ^{32}P-labelled base-cleaved post-nuclear poly(A)-containing RNA was incubated under stringent hybridization conditions with each of the four plasmid DNAs immobilized on paper as described above. Hybridized ^{32}P-labelled mRNA sequences were recovered, and each then used as a sequence-specific hybridization probe *in situ*, to determine the distribution of the individual milk protein mRNA sequences within our clone-bank. The 131 colonies which had previously hybridized to total ^{32}P-labelled post-nuclear poly(A)-containing RNA were replica

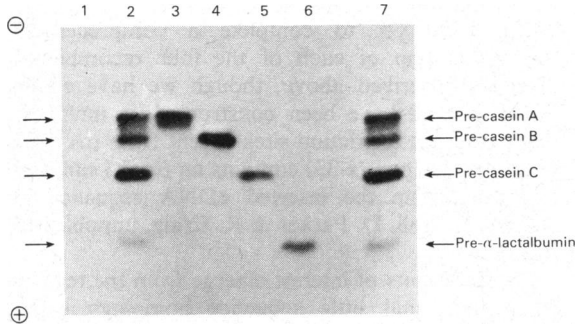

Fig. 3. *Identification of recombinant plasmids containing milk protein cDNA sequences*

Sequence-specific poly(A)-containing RNA isolated by hybridization to partially restricted, denatured recombinant DNA immobilized on DBM-paper was added to a nuclease-treated reticulocyte lysate, and the [^{35}S]methionine-labelled cell-free translation products were separated by SDS/polyacrylamide-gel electrophoresis, then visualized by fluorography (see the Materials and methods section). Lane 1, no added RNA; lane 2, lactating mammary gland post-nuclear poly(A)-containing RNA; lanes 3–6, RNA isolated by hybridization to plasmids pgp N-33, pgp K-27, pgp O-22, and pgp K-9 respectively; lane 7, as lane 2. Arrows denote the primary translation products (Craig *et al.*, 1978*b*).

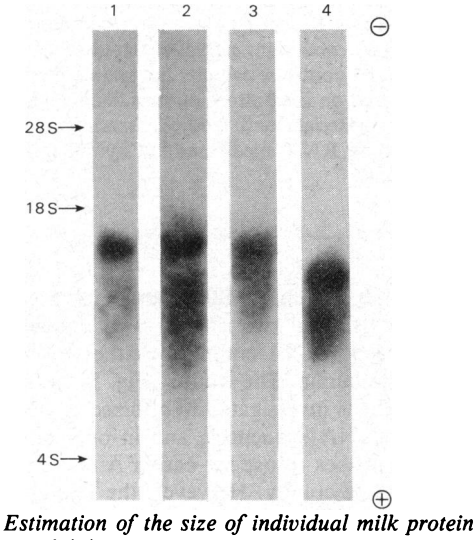

Fig. 4. *Estimation of the size of individual milk protein poly(A)-containing mRNA sequences*

Total cellular poly(A)-containing RNA was isolated from the lactating guinea-pig mammary gland as described by Hall *et al.* (1979). Four 1 μg amounts were then treated with glyoxal, loaded onto separate slots on a 1.5% (w/v) agarose gel, electrophoresed and then blotted onto DBM-paper as described in the Materials and methods section. This was then cut into strips, and the position of the four milk protein mRNA sequences was determined by hybridization to ^{32}P-nick-translated sequence-specific recombinant plasmid (see the Materials and methods section). Lane 1, pgp N-33 (casein A); lane 2, pgp K-27 (casein B); lane 3, pgp O-22 (casein C); lane 4, pgp K-9 (α-lactalbumin). Size estimation of the individual mRNA sequences was determined by direct comparison with the mobilities of glyoxal-treated reticulocyte 28S (5300 bases) and 18S (2060 bases) ribosomal RNA, and *E. coli* tRNA (4S; 80 bases) marker species that were electrophoresed on parallel tracks, and whose positions were visualized by staining in Acridine Orange (McMaster & Carmichael, 1977).

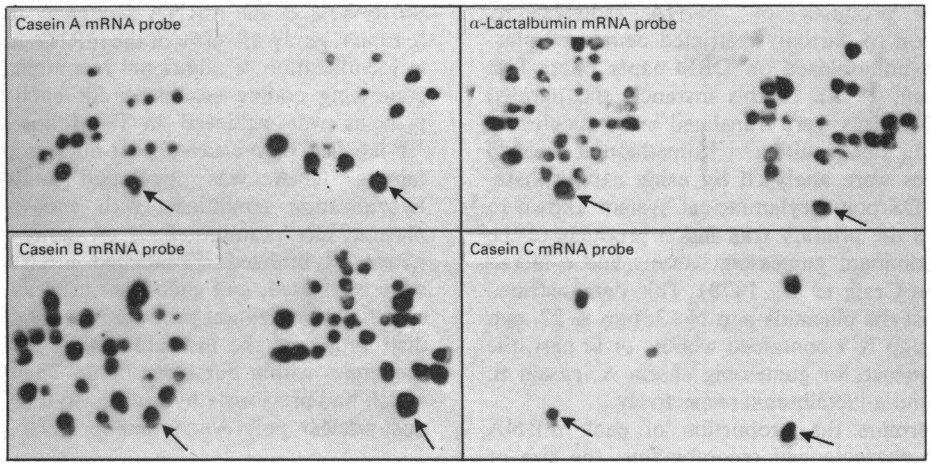

Fig. 5. *Analysis of the distribution of individual cloned milk protein cDNA sequences within the cDNA library*
All colonies previously shown by hybridization *in situ* with total ^{32}P-labelled post-nuclear poly(A)-containing RNA to contain abundant polyadenylated RNA sequences were replica plated onto Millipore filters (four sets of two filters). As controls, colonies containing the characterized plasmids pgp N-33, pgp K-27, pgp O-22 and pgp K-9 were duplicated on the bottom of each filter. Hybridization *in situ* was then performed with each set of filters as described in the Materials and methods section by using a different sequence-specific ^{32}P-labelled base-cleaved poly(A)-containing RNA species as the hybridization probe in each instance. Arrows denote the position of the control plasmids.

plated onto Millipore filters, and each set (four sets each containing two filters) was incubated with ^{32}P-labelled mRNA coding for either casein A, B, C or α-lactalbumin. The results (Fig. 5) show that 117 of the clones investigated hybridized to one or other of the mRNA sequences, and in only one instance did two probes (those for casein A and B) hybridize to the same colony. However, the relative distribution of the different milk protein mRNA sequences was unexpected. Excluding the single colony mentioned above, of the colonies found to contain milk protein mRNA sequences, 22 (19%) contained casein A mRNA sequences, 45 (39%) casein B and 46 (40%) α-lactalbumin. Only three (2%) contained casein C mRNA sequences.

Discussion

Our cloning strategy was based on our earlier observation that 55% of the post-nuclear poly(A)-containing RNA isolated from the lactating guinea-pig mammary gland is comprised of milk protein mRNA sequences (Craig *et al.*, 1979). The great abundance of these sequences obviates the requirement to purify the individual milk protein mRNA sequences even for the preliminary screening of our library of colonies by hybridization *in situ*. The host/vector system and site of insertion were selected so as to minimize the likelihood of peptide expression, and so permit characterized recombinants, shown to contain milk protein cDNA sequences inserted at the *Eco*RI site, to be manipulated in rec A$^-$ host cells under conditions of good microbiological practice.

We have yet to complete a comprehensive restriction map of each of the four recombinant plasmids described above, though we have established that all have been constructed as intended. Each contains restriction sites absent from pAT153 but only one (pgp N-33) contains an *Eco*RI site, and this lies within the inserted cDNA sequence (J Allison, L. Hall, D. Parker & R. Craig, unpublished work).

Several points of interest emerge from the results. It is clear that little sequence homology exists between any of the caseins or α-lactalbumin, since the hybridization conditions used for the positive hybridization–translation procedure result in the clean separation of the individual milk protein mRNA sequences, as judged by cell-free protein synthesis and SDS/polyacrylamide-gel electrophoresis in a single dimension. This procedure did, however, produce evidence for heterogeneity within the primary translation product of casein A (pre-casein A). Subsequent detailed analysis of all four primary translation products by two-dimensional polyacrylamide-gel electrophoresis confirms this observation, and furthermore demonstrates considerable heterogeneity within pre-casein A and pre-casein B (see Pascall *et al.*, 1981).

Comparison of the sizes of the inserted sequences with that of the poly(A)-containing milk protein mRNA demonstrates that, with the exception of casein C, our procedures have generated recombinant plasmids containing large, if not full-length, copies of the milk protein mRNA sequences. Moreover, as expected for an abundant mRNA population (Craig *et al.*, 1979), these represent about 35% of the total recombinant colonies obtained. (This calculation makes allowance for the background of unrestricted pAT153.)

Casein C mRNA is equally abundant *in vivo* (Burditt *et al.*, 1981), but plasmids containing casein C cDNA sequences were present at a very low frequency (<1% of the total recombinants), and those identified contained at best 60% of the mRNA sequence. Amino acid analysis of casein C (Craig *et al.*, 1978*a*) demonstrates that, when compared with caseins A and B, or α-lactalbumin, it contains exceptionally high amounts of proline, alanine, arginine, valine and leucine. The codons for the first three amino acids, independent of the codon usage, are rich in G + C, whereas the remainder, depending on the codon usage, may also be considered to be G + C enriched, particularly as there appears to be a preference for C or G in the third position of eukaryotic codons (see Shine *et al.*, 1977). On the basis of these observations it seems reasonable to predict that an mRNA of high G + C content will contain stable secondary structure. This might well restrict the free movement of the avian-myeloblastosis-virus reverse transcriptase along the template strand (see Payvar & Schimke, 1979) resulting in low yields and small inserted sequences in the resulting recombinants.

Although casein C, and the primary translation product pre-casein C, appear to be of considerably lower molecular weight than caseins A and B as judged by SDS/polyacrylamide-gel electrophoresis, all the caseins appear to be encoded by mRNAs of similar size after treatment with glyoxal. We have independently verified, by separation of the milk protein mRNA species on denaturing sucrose gradients in the presence of formamide, that caseins A and B are encoded by mRNAs of similar size. However, under these conditions, casein C mRNA was significantly larger (D. Parker & R. Craig, unpublished work). These unexpected observations may in part reflect the amino acid composition of casein C, which contains a high percentage of low-molecular-weight residues. Consequently, when compared with casein B (see Craig *et al.*, 1978*a*), casein C has a lower molecular weight, but contains 13 additional amino acid residues, and therefore will require additional coding capacity within the mRNA sequence. It is also possible that the predicted high G + C content of casein C mRNA may give rise to regions of stable secondary structure even in 70% formamide, resulting in an increased sedimentation rate.

In addition to our work on the guinea-pig, we have recently constructed a library of cDNA recombinants from lactating human mammary gland poly(A)-containing RNA, and have isolated and characterized recombinants containing the α-lactalbumin cDNA sequence (Hall *et al.*, 1981). Others have recently cloned rat casein and α-lactalbumin cDNA sequences (Richards *et al.*, 1980*a*,*b*), whereas recombinants containing milk protein cDNA sequences from several other species are at present under construction in other laboratories. Detailed DNA sequence analysis of all these genes should provide interesting insights into the evolution of milk proteins among the different species. Moreover, the availability of cloned sequences will now (i) allow analysis of the structure and organization of the milk protein genomic DNA sequences, (ii) provide a source of sequence-specific hybridization probes for kinetic studies on mRNA synthesis and maturation, and the effect of hormones on these mechanisms (see Guyette *et al.*, 1979), and (iii) as we describe elsewhere (Burditt *et al.*, 1981; Pascall *et al.*, 1981) provide the means of rapidly isolating individual mRNA or cDNA sequences from complex populations, for use in cell-free protein synthesis or nucleic acid hybridization studies.

We thank Miss J. Allison for skilled technical assistance, and Dr. J. F. Williams for technical advice and many invaluable discussions. pAT153 DNA was the gift of Professor D. Sherratt, and avian-myeloblastosis-virus reverse transcriptase was generously provided by Dr. J. Beard and Dr. J. Gruber. Terminal deoxynucleotidyl transferase and *Hpa*II were the gifts of Dr. I. R. Johnston and Dr. B. Griffin respectively. We are also grateful to the Wellcome Trust for their generous contribution towards the equipment in our containment facility.

References

Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5350–5354

Banerjee, M. R. (1976) *Int. Rev. Cytol.* **47**, 1–97

Bathurst, I. C., Craig, R. K., Herries, D. G. & Campbell, P. N. (1980*a*) *Eur. J. Biochem.* **109**, 183–191

Bathurst, I. C., Craig, R. K., Herries, D. G. & Campbell, P. N. (1980*b*) *Biochem. J.* **192**, 489–498

Bonner, W. M. & Laskey, R. A. (1974) *Eur. J. Biochem.* **46**, 83–88

Boseley, P. G., Tuyns, A. & Birnstiel, M. L. (1978) *Nucleic Acids Res.* **5**, 1121–1137

Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472

Burckhardt, J., Telford, J. & Birnstiel, M. L. (1979) *Nucleic Acids Res.* **6**, 2963–2971

Burditt, L. J., Parker, D., Craig, R. K., Getova, T. & Campbell, P. N. (1981) *Biochem. J.* **194**, 999–1006

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*a*) *Biochem. J.* **173**, 633–641

Craig, R., Boulton, A., Campbell, P. N., Lane, C., Mellor, A. & Perrera, P. (1978*b*) *FEBS Symp.* **53**, 43–55

Craig, R. K., Boulton, A. P., Harrison, O. S., Parker, D. & Campbell, P. N. (1979) *Biochem. J.* **181**, 737–756

Craig, R. K., Bathurst, I. C. & Herries, D. G. (1980) *Nature (London)* **288**, 618–619

Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961–3965

Guyette, W. A., Matusik, R. J. & Rosen, J. M. (1979) *Cell* **17**, 1013–1023

Hall, L., Craig, R. K. & Campbell, P. N. (1979) *Nature (London)* **277**, 54–56

Hall, L., Davies, M. S. & Craig, R. K. (1981) *Nucleic Acids Res.* **9**, 65–84

Hofstetter, H., Schambock, A., Van Den Berg, J. & Weissmann, C. (1976) *Biochim. Biophys. Acta* **454**, 587–591

Katz, L., Kingbury, D. T. & Helinski, D. R. (1973) *J. Bacteriol.* **114**, 577–591

Kay, R. M., Harris, R., Patient, R. K. & Williams, J. G. (1980) *Nucleic Acids Res.* **8**, 2691–2707

McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146

McMaster, G. K. & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4835–4838

McReynolds, L. A., Catterall, J. F. & O'Malley, B. W. (1977) *Gene* **2**, 217–231

Pascall, J. C., Boulton, A. P., Parker, D., Hall, L. & Craig, R. K. (1981) *Biochem. J.*, in the press

Payvar, F. & Schimke, R. T. (1979) *J. Biol. Chem.* **254**, 7636–7642

Pelham, H. R. B. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67**, 247–256

Richards, D. A., Rodgers, J. R., Supowit, S. C. & Rosen, J. M. (1980*a*) *J. Biol. Chem.*, in the press

Richards, D. A., Blackburn, D. E. & Rosen, J. M. (1980*b*) *J. Biol. Chem.*, in the press

Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251

Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. (1977) *Nature (London)* **270**, 494–499

Smith, D. F., Searle, P. F. & Williams, J. G. (1979) *Nucleic Acids Res.* **6**, 487–506

Sutcliffe, J. G. (1978*a*) *Cold Spring Harb. Symp. Quant. Biol.* **43**, 77–90

Sutcliffe, J. G. (1978*b*) *Nucleic Acids Res.* **5**, 2721–2728

Twigg, A. J. & Sherratt, D. (1980) *Nature (London)* **283**, 216–218

Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3683–3687

Zimmerman, S. B. & Sandeen, G. (1966) *Anal. Biochem.* **14**, 269–277