# 5'-Terminal sequences of eucaryotic mRNA can be cloned with high efficiency

Hartmut Land, Manuel Grez, Hansjörg Hauser+, Werner Lindenmaier+ and Günther Schütz

Institut für Zell- und Tumorbiologie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-6900 Heidelberg, GFR

Received 27 March 1981

## ABSTRACT

A method for cloning mRNAs has been used which results in a high yield of recombinants containing complete 5'-terminal mRNA sequences. It is not dependent on self-priming to generate double-stranded DNA and therefore the S1 nuclease digestion step is not required. Instead, the cDNA is dCMP-tailed at its 3'-end with terminal deoxynucleotidyl transferase (TdT). The synthesis of the second strand is primed by oligo(dG) hybridized to the 3'-tail. Double-stranded cDNA is subsequently tailed with dCTP and annealed to dGMP-tailed vector DNA. This approach overcomes the loss of the 5'-terminal mRNA sequences and the problem of artifacts which may be introduced into cloned cDNA sequences.

Chicken lysozyme cDNA was cloned into pBR322 by this procedure with a transformation efficiency of 5 x $10^3$ recombinant clones per ng of ds-cDNA. Sequence analysis revealed that at least nine out of nineteen randomly isolated plasmids contained the entire 5'-untranslated mRNA sequence. The data strongly support the conclusion that the 5'-untranslated region of the lysozyme mRNA is heterogeneous in length.

## INTRODUCTION

A variety of methods for cloning eucaryotic mRNA sequences has been described (1-6). In the procedure most commonly used (4-6), the ability of single-stranded cDNA to form a hairpin at its 3'-end is exploited to prime the synthesis of the second DNA strand. Before inserting this double-stranded cDNA into a plasmid, it is necessary to digest the single-stranded hairpin loop of the ds-cDNA with S1 nuclease. This invariably results in the loss of sequences corresponding to the extreme 5'-terminal region of the mRNA (e.g. 4,7-10).

Cloning of the entire 5'-untranslated region of a mRNA is, however, highly desirable. The presence of the entire 5'-end would facilitate mapping of the mRNA onto the corresponding genomic DNA, particularly with respect to the start site of transcription (11), and it may also in-

crease the probability of expressing cloned structural genes in bacteria (12).

By inserting cDNA-mRNA hybrids into a plasmid vector, Zain et al. (13) were able to clone a complete copy of the leader segments of adenovirus 2 fiber mRNA. Because this technique gives only low yields of recombinants, we have chosen a procedure similar to the one used by Rougeon et al. (1) and Cooke et al. (14), in which S1 nuclease digestion of double-stranded cDNA is avoided for construction of recombinant plasmids.

Poly(A)-containing RNA was reverse-transcribed by AMV reverse transcriptase under conditions which increase the yield of full length cDNA. After alkaline hydrolysis of the mRNA the 3'-end of the cDNA was dCMP-tailed with terminal deoxynucleotidyl transferase. The second cDNA strand synthesis was primed by oligo$(dG)_{12-18}$ hybridized to the 3'-homopolymer tail. Full length ds-cDNA copies were purified on a preparative agarose gel. After a second tailing step with dCTP the cDNA was annealed to plasmid DNA tailed with dGTP. The hybrid DNA was transformed and amplified in bacteria. The protocol is summarized in Figure 1.

## MATERIALS AND METHODS

### 1. Enzymes

Avian myeloblastosis virus reverse transcriptase was kindly provided by Dr. J.W. Beard through the Office of Program Resources and Logistics, National Cancer Institute of the U.S.A. Terminal deoxyribonucleotidyl transferase was obtained from Bethesda Research Laboratories, Neu-Isenburg; endonuclease S1 from *Aspergillus oryzae* was Type III enzyme from Sigma, Munich; DNA polymerase I, the Klenow fragment of DNA polymerase I, alkaline phosphatase grade 1 from calf intestine, and T4 polynucleotide kinase were from Boehringer, Mannheim. Restriction endonucleases were obtained from Boehringer, Mannheim or from New England Biolabs, Beverly; *Pst*I was from Renner, Mutterstadt.

### 2. Synthesis of cDNA

Polysomal poly$(A)^+$ RNA from hen oviduct cells was prepared as described (15). Reverse transcription of total poly$(A)^+$ RNA (max. 100 µg/ml) (15) was carried out in 50 mM Tris-HCl, pH 8.3, 50 mM KCl, 8 mM $MgCl_2$, 0.4 mM DTT, 30 µg/ml oligo-dT, 0.1 mg/ml actinomycin D, 1 mM dGTP, dCTP and dTTP each and 0.2 mM $\alpha[^{32}P]$ ATP (0.75 Ci/mmole = 27.75 GBq) for 2 min at room temperature followed by incubation at 42°C for 60 min. Reverse transcriptase was

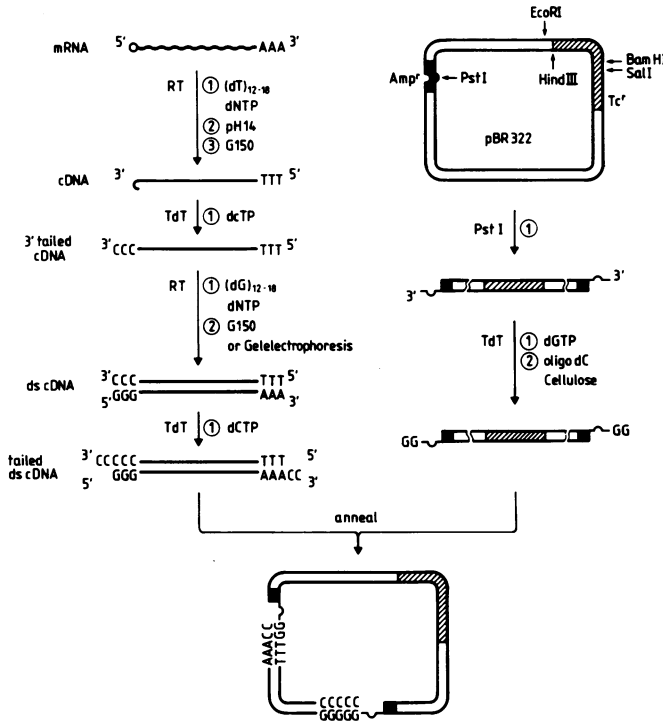


Fig. 1 Schematic diagram of the construction of recombinant plasmids containing eukaryotic mRNA sequences. Conditions which conserve the 5'-terminal sequences of mRNA are shown.
Numbers indicate separate biochemical steps between the re-isolation of the products by ethanol precipitation. RT, reverse transcriptase; TdT, terminal deoxynucleotidyl transferase; $Amp^r$, β-lactamase gene; $Tc^r$, tetracycline resistance gene.

used at a concentration of 5 U/µg mRNA. The reaction was stopped by adjusting the mixture to 0.5 M NaCl, 20 mM EDTA, 0.2 % SDS. After extraction with chloroform/isoamylalcohol (24:1 v/v) the aqueous phase was brought to 0.4 N NaOH and left at 25°C overnight. Nucleotides and oligodeoxynucleotides were separated from the cDNA by gel-filtration through Sephadex G150 in 5 mM Tris-HCl, pH 7.5, 5 mM NaCl. The cDNA was precipitated with ethanol or lyophylized without using nucleic acid carrier.

3. Tailing of the cDNA with terminal deoxyribonucleotidyl transferase (TdT)

1 to 5 pmoles of single-stranded cDNA were tailed at the 3'-ends with terminal deoxyribonucleotidyl transferase (TdT) (16) in a reaction con-

taining 140 mM cacodylic acid (free acid/Sigma), 60 mM Tris base adjusted with KOH to a final pH of 7.6, 1 mM $CoCl_2$, 0.1 mM DTT and 0.1 mM [$^3$H]dCTP (200-2000 cpm/pmole) in a volume of 100 µl. After preincubation for 5 min at 37°C and chilling in ice the reaction was started with 7.5 U of TdT per 1 pmole of cDNA and kept for 2-5 min at 15°C. The reaction mixture was heated for 5 min at 70°C in 0.5 M NaCl, 10 mM EDTA, extracted with chloroform/isoamylalcohol, and the cDNA was ethanol precipitated twice.

4. Synthesis of ds-cDNA on 3'-dCMP-tailed cDNA

Oligo$(dG)_{12-18}$ (Collaborative Res.) was hybridized to the 3'-dCMP-tailed cDNA to prime the synthesis of the second strand. Hybridization was carried out in 50 mM Tris-HCl, pH 8.3, 30 mM KCl, 10 mM $MgCl_2$, 30 µg/ml oligo(dG). The temperature was kept at 68°C and 55°C for 5 min each, at 50°C for 10 min and at 43°C for 15 min. After chilling on ice, the reaction mixture containing cDNA concentrations of up to 10 µg/ml was adjusted to 10 mM DTT, 1 mM of each dNTP and was incubated with 50-60 U of reverse transcriptase per µg of cDNA for 10 min at 37°C and for 60 min at 42°C. The reaction was stopped as described above. Ds-cDNA was further purified by gel-filtration through G150 or was run on a 2 % agarose gel and the desired size class was cut out. Ds-cDNA was recovered from agarose by the glass binding method as described by Gillespie and Vogelstein (17).

Ds-cDNA was tailed under the same conditions as described above, except that the reaction was incubated for 1 min at 37°C.

5. Analysis of cDNA size

Linearized plasmid DNAs pls-1 (7), pom-48 (18), and poa-1210 (H. Land, unpublished) were covalently coupled to DBM-cellulose papers (5 µg DNA/cm$^2$) (19). 125 ng of cDNA or 300 ng of ds-cDNA were hybridized to these plasmids (150 µg each) in 2.5 ml of 50 % formamide, 4 x SSC, 0.1 % SDS, 1 mM EDTA, 0.5 mg/ml salmon sperm DNA and 0.1 mg/ml poly(A). Hybridization was carried out for 18 hours at 42°C. Papers were washed four times for 30 min in 0.2 x SSC, 0.1 % SDS at 65°C. Elution was performed at 65°C in 99 % formamide, 10 mM Tris-HCl, pH 7.5, 0.1 % SDS, 1 mM EDTA and 7.5 µg/ml salmon sperm DNA. Hybridized and subsequently eluted cDNA was ethanol precipitated and run on a 2 % alkaline agarose gel at 80 mA in 30 mM NaOH, 2 mM EDTA.

6. Construction of recombinant plasmid DNA

Plasmid pBR322 was cut with PstI and dGMP-tailed with TdT as described above. About 10 dGMP residues were added to each 3'-end (20). To remove

non-tailed molecules, the DNA was passed over an oligo-dC cellulose column. DNA prepared in this manner has about 0.02 % of the transformation efficiency when compared to supercoiled pBR322. Tailed pBR322 was a gift from W. Roewekamp.

Tailed pBR322 and tailed ds-cDNA were mixed in a molar ratio of 1:1 with a final concentration of 250 ng vector/ml in 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA. The mixture was incubated first at 68°C for 5 min and then at 43°C for 2 hr, and cooled down to room temperature during a 2 hr period. 10 µl of annealing solution (2.5 ng vector) were taken for 100 µl of competent cells. Transformation of E.coli 5K (21) was carried out exactly as described by Dagert and Ehrlich (22). The transformation efficiency of E.coli 5K was 1-2 x $10^7$ transformants/µg of supercoiled pBR322. 1 µg of annealed vector DNA gave rise to about 8 x $10^5$ tetracycline resistant transformants. About 99 % of the transformants contained cDNA sequences.

7. Screening for plasmids containing lysozyme mRNA sequences

Colony screening was done as described previously (7). Nick translated insert DNA of lysozyme cDNA plasmid pls-1 (7) (3 x $10^7$ cpm/µg) was used to probe for plasmids containing lysozyme mRNA sequences.

8. Sequencing

DNA sequence analysis was done as described by Seif et al. (23). mRNA sequencing using reverse transcriptase was carried out as described by Schreier and Cortese (24). Primer DNA fragments were obtained from plasmid pls-1.

9. Biosafety conditions

Construction and growth of recombinant plasmids were conducted under L3/B1 conditions as specified by the Zentrale Kommission für Biologische Sicherheit of the Federal Republic of Germany.

## RESULTS

1. Synthesis of full length cDNA

To clone the entire mRNA sequence including the 5'-terminus in a bacterial plasmid, it is a prerequisite to synthesize a full length DNA copy of the mRNA. Optimal conditions were therefore determined (15) which

allow the synthesis of full length cDNAs. cDNA was synthesized from chicken oviduct poly(A)-containing RNA as described under methods (Fig. 2 lane a). To determine the lengths of specific cDNAs, total cDNA was hybridized to plasmid DNA bound to DBM-cellulose paper. DNA of the plasmids pls-1 (lysozyme), pom-48 (ovomucoid) and poa-1210 (ovalbumin) was used. After extensive washing of the papers, the hybridized cDNAs were eluted and run on a 2 % alkaline agarose gel (Fig. 2, lanes c,d,e). cDNA copies with a length as expected from the size of the corresponding mRNAs (25-27) can be recognized for lysozyme (Fig. 2, lane c), ovomucoid (lane d) and ovalbumin cDNA (lane e). As determined from the intensity of the bands approximately 75 % of the lysozyme cDNA molecules appeared as full size copies. The proportion of full length cDNA molecules decreased with increasing size of

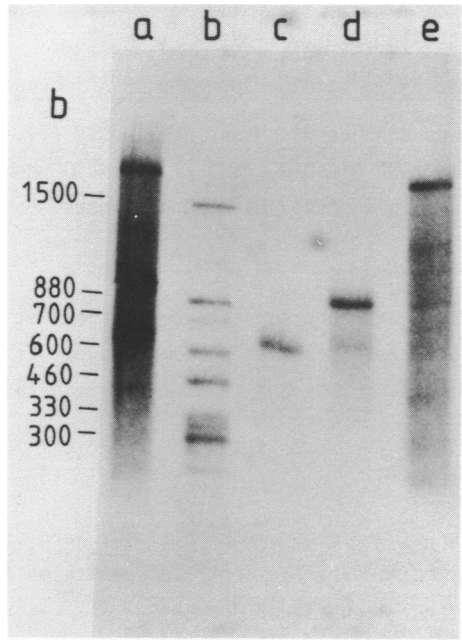


Fig. 2 Size of cDNAs synthesized from chicken oviduct polysomal poly(A)$^+$ RNA.
DNA was electrophoresed on a 2 % alkaline agarose gel. The gel was dried and autoradiographed. Lane a) cDNA from chicken oviduct poly(A)$^+$ RNA; b) labelled λdv-1 DNA cleaved with HinfI. The numbers indicate the length of DNA fragments in bases (b). cDNAs specific for lysozyme-mRNA (c), ovomucoid-mRNA (d) and ovalbumin-mRNA (e) were purified by hybridization to cDNA plasmids covalently coupled to DBM cellulose paper. For experimental details see text.

the mRNA templates, with ovomucoid mRNA yielding about 50 % and ovalbumin mRNA about 5 % of full sized cDNA copies.

2. Synthesis of double-stranded cDNA

The most commonly used method of preparing double-stranded cDNA relies on the property of single-stranded cDNA to form hairpin-like structures at its 3'-end. These loops can function as primers for the synthesis of the second DNA strand. However, since the hairpin structures have to be digested with S1 nuclease, the 5'-end sequences of the mRNA are always lost. Tailing the cDNA at the 3'-end with dCTP and priming the synthesis of the second strand with oligo$(dG)_{12-18}$ avoids the covalent linkage of both DNA strands and thus results in the conservation of the sequences corresponding to the 5'-terminus of the mRNA.

In the procedure described here, the conservation of the 5'-terminal mRNA sequences is strictly dependent on the purity of the available terminal transferase. cDNA derived from chicken poly$(A)^+$-containing RNA was run on a 2 % alkaline agarose gel before and after incubation with terminal transferase. Degradation of the cDNA during the tailing reaction was not observed under the conditions used (Fig. 3, lanes a and b). As estimated from the incorporation of radioactivity and from the increase in size (Fig. 3, lane b), about 20-25 dCMP residues were added to the 3'-ends of the cDNA. An approximately two-fold molar excess of enzyme over DNA was chosen to assure a quantitative tailing of the DNA molecules.

In order to test the efficiency of 3'-tailed cDNA in directing the synthesis of the second strand, reactions were carried out in the presence and the absence of a complementary primer. Synthesis of the second strand was followed by determination of the amount of radioactive cDNA that was rendered S1 nuclease resistant. In all experiments 70-80 % of the cDNA was protected when tailed cDNA was used in combination with the oligonucleotide-primer. In the absence of primer about 30 % of the cDNA became resistant to S1 nuclease. In comparison, second strand synthesis with non-tailed cDNA as template led to 40-60 % S1 nuclease resistant cDNA (for a typical experiment see Table 1).

Therefore, the presence of oligo(dG) primer leads to an increased yield of double-stranded cDNA. The size of the second DNA strand synthesized from the 3'-tailed cDNA template was revealed by following the length of the radioactive cDNA after treatment of the reaction products with S1 nuclease (Fig. 3, lane d,e). In an experiment similar to the one shown in Fig. 2

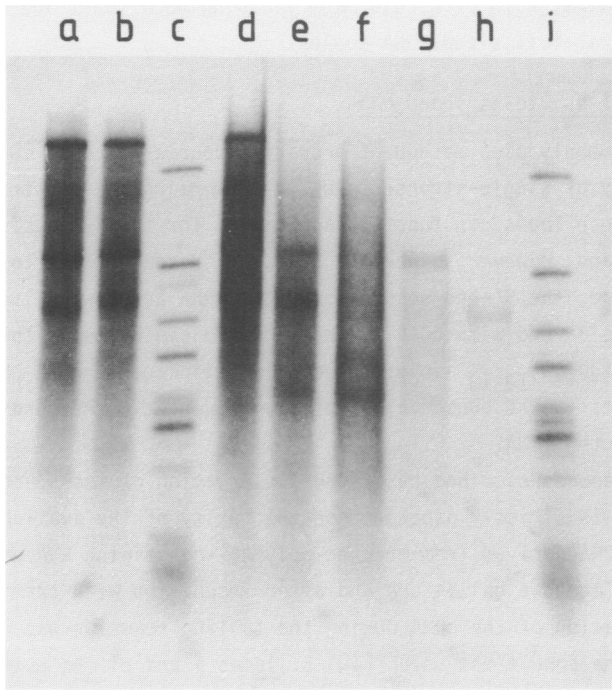


Fig. 3 Size of double-stranded cDNA synthesized from cDNA tailed at its 3'-end.
DNA was electrophoresed on a 2 % alkaline agarose gel. The gel was dried and autoradiographed. Lane a) cDNA from chicken oviduct poly(A)$^+$ RNA; b) cDNA after tailing at the 3'-end using terminal deoxynucleotidyl transferase (TdT), c) and i) labelled marker, λdv-1 DNA cleaved with HinfI. The length of the DNA fragments is indicated in Fig. 1. d) Double-stranded cDNA (ds-cDNA) synthesized as described with tailed cDNA shown in (b) as template. Only the cDNA is labelled. The size of the second DNA strand was analyzed by following the radioacitive cDNA remaining resistant after S1 nuclease digestion of the ds-cDNA (lane e). In lanes f-h the same kind of analysis is shown after hybridizing S1 nuclease digested ds-cDNA to plasmid DNA specific for ovalbumin, ovomucoid and lysozyme cDNA respectively, which was bound to DBM cellulose paper. For experimental details see text.

(lanes c-e), total ds-cDNA was digested with S1 nuclease, denatured and hybridized to plasmid DNA bound to DBM-cellulose paper. Hybridized DNAs specific for ovalbumin, ovomucoid and lysozyme cDNA sequences were eluted and run on a 2 % alkaline agarose gel (Fig. 3, lanes f,g,h, respectively). As observed with the synthesis of the first strand, the proportion of full-length ds-cDNA molecules decreased with increasing template length. About 30 % of the lysozyme ds-cDNA molecules respresented full length copies.

Table 1: Synthesis of double-stranded cDNA

| | -S1 nuclease (cpm) | +S1 nuclease (cpm) | cDNA which is S1 nuclease resistant (%) |
|---|---|---|---|
| Tailed cDNA alone | 1318 | 203 | 15 |
| Tailed cDNA -oligo(dG)$_{12-18}$ +RT | 1320 | 463 | 35 |
| Tailed cDNA +oligo(dG)$_{12-18}$ +RT | 1910 | 1470 | 77 |
| cDNA +RT | 1480 | 829 | 56 |

Radioactive cDNA was used as template for the synthesis of ds-cDNA (see methods). The second DNA strand was not labelled. Synthesis of the second strand was followed by determining the amount of radioactive cDNA becoming resistant to S1 nuclease digestion. Aliquots were incubated in 30 mM sodium acetate pH 4.5, 3 mM $ZnSO_4$, 0.3 M NaCl and 10 µg/ml salmon sperm DNA at 37°C for 60 min. 250 U/ml of S1 nuclease were used. RT, reverse transcriptase.

After synthesis of the second strand the ds-cDNA was extracted, passed over a G150 column and tailed with dCTP to generate 7-10 b long dCMP-tails at each 3'-end (20).

3. Bacterial transformation with hybrid plasmid DNA

pBR322 was cut with PstI and tailed with dGTP. 5 ng were annealed to 0.7 ng of unsized dCMP-tailed ds-cDNA. After transformation (22) into E.coli 5K about 4000 tetracycline resistant colonies were obtained. From 1800 transformants screened, 90 (5 %) contained lysozyme cDNA plasmids (see Methods). This yield of recombinants reflects the abundance of lysozyme mRNA among chicken oviduct poly(A)-containing RNA (2-3 %) (15).

4. Lysozyme cDNA plasmids contain the entire mRNA 5'-sequence

Plasmid DNA from twenty-three randomly taken lysozyme cDNA clones was prepared (28). From nineteen plasmids the insert DNA could be excised by digestion with PstI, since the recognition sequence was restored by the G-C tailing. All plasmids contained at least one PstI site. Two plasmids carrying inserts of about 650 bp in length were assumed to carry a full

length cDNA copy of lysozyme mRNA (620-630 b). The other inserts were heterogeneous in size (see Table 2).

Restriction endonuclease analysis revealed that a minimum of nine out of a total of nineteen plasmids may contain the sequence corresponding to the entire 5'-end of the mRNA (data not shown). These nine plasmids were sequenced in the region of interest. The sequencing strategy is shown in Fig. 4a. The plasmid DNA was cleaved with BstNI, labelled at the 5'-ends with $\gamma[^{32}P]ATP$ and run on a 5 % polyacrylamide gel. The appropriate DNA fragment was isolated and cut with HpaII. This method was applicable for both orientations of the insert in the plasmid. Fig. 4b shows sequences of two of these plasmids, pls-184 and pls-582, which correspond to the 5'-end of the lysozyme mRNA, the GC-tail and sequences of pBR322 flanking the PstI site. The end of the lysozyme cDNA sequence differs by two nucleotides. S1 nuclease mapping of the mRNA (M. Grez et al., submitted) and direct sequencing of the capped 5'-termini (R.E. Rhoads, personal communication) have revealed two lysozyme mRNAs, whose 5'-end-sequences are identical to those found in the cloned cDNAs, mapping 2 nucleotides apart. In 5 of the 9 plasmids sequenced the mRNA sequence stops at the site shown for pls-184. The other 4 plasmids carry the same cDNA as pls-582. These results are

Table 2: Collection of lysozyme cDNA clones containing the entire 5'-terminal mRNA sequence.

| Plasmid | Length of the insert in bp | Orientation of the mRNA sequence | Length of the GC-tail at the 5'-end in bp | 5'-End of the lysozyme RNA sequence |
|---|---|---|---|---|
| pls- 136 | 460 | A | 25 | tail G/GCAGTCCCG |
| pls- 157 | 520 | A | 39 | G/GCAGTCCCG |
| pls- 184 | 650 | A | 28 | G/GCAGTCCCG |
| pls- 582 | 630 | A | 19 | G/ AGTCCCG |
| pls- 687 | 550 | B | 37 | G/GCAGTCCCG |
| pls- 706 | 420 | A | 33 | G/GCAGTCCCG |
| pls- 713 | 330 | N.D. | N.D. | G/ AGTCCCG |
| pls- 739 | 600 | B | N.D. | G/ AGTCCCG |
| pls-1023 | 650 | B | 29 | G/ AGTCCCG |
| pls-1124 | | | tail G/AGAGGCAGGTGCAAGAGAGCTTGCAGTCCCG | |
| Sequence of the genomic DNA | | | GAAGTTAAAAGAAGAGGCAGGTGCAAGAGAGCTTGCAGTCCCG<br>-24 -2+1 | |

The orientation is explained in Fig. 4. The sequence of the genomic DNA at the 5'-end of the lysozyme gene (M. Grez et al., submitted) is given for comparison with the three groups of lysozyme cDNA clones. The numbers indicate the coordinates of the mRNA start positions on the genomic DNA. The sequence of the non-coding strand is shown.

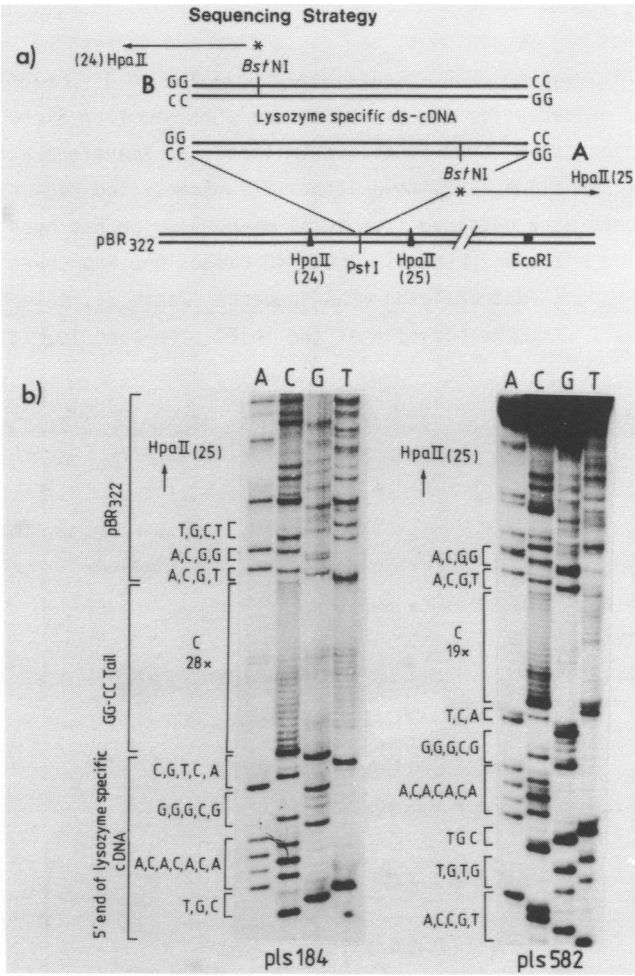


Fig. 4: Sequencing strategy and DNA sequence analysis of two plasmids containing lysozyme mRNA sequences
(a) Sequencing strategy. A and B indicate the two possible orientations of the cDNA insert in pBR322.
(b) DNA sequence analysis of pls-184 and pls-582 showing the region corresponding to the 5'-untranslated sequence of the lysozyme mRNA. pls-184 and pls-582 contain 31 or 29 nucleotides upstream from the AUG codon of the lysozyme mRNA, respectively. The sequence of the coding strand is shown.

summarized in Table 2.

The sequence of the 5'-untranslated region of the lysozyme mRNA which was obtained by reverse transcription under chain termination conditions

(24) (Fig. 5) led to three distinct stop-bands (arrows in Fig. 5). The two major bands are separated by one nucleotide and correspond exactly to the existence of a lysozyme mRNA species carrying a longer 5' non-coding region. Further evidence for its existence was obtained from S1 nuclease mapping experiments (M. Grez et al., submitted). We therefore attempted to isolate a cDNA clone carrying a longer 5'-untranslated region. Lysozyme cDNA clones were obtained from total ds-cDNA which had been size fractionated on a 2 % agarose gel. Among 40 clones one was shown to contain a 53 bp long 5'-untranslated mRNA sequence (Table 2), a length which corresponds well with the position of the third stop-band in Fig. 5.

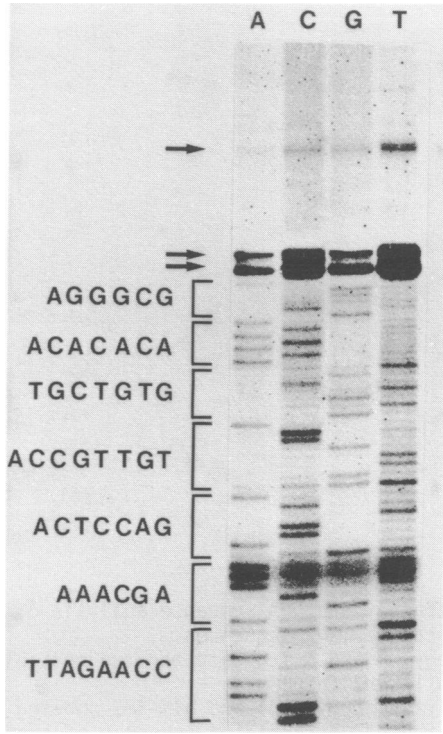


Fig. 5: Reverse transcription of lysozyme mRNA under chain termination conditions.
Polysomal poly(A)$^+$ RNA extracted from hen oviduct cells was hybridized to a 2-fold molar excess of an exonuclease III treated BamH I-Alu I DNA fragment obtained from lysozyme cDNA clone pls-1 (7). Primer extension was carried out as described (24). Arrows indicate the major stop bands observed upon reverse transcription of the lysozyme mRNA. The DNA sequence corresponds to the coding strand.

## DISCUSSION

The experiments described here show that it is possible to clone the entire 5'-terminal sequence of a mRNA in a bacterial plasmid with high efficiency. cDNA probes and structural gene sequences derived from these clones facilitate the mapping of mRNA start-sites, which are also presumed to be the sites for initiation of transcription (29-32). The method presented here may also increase the chance of synthesizing structural gene sequences suitable for expression in bacteria (12).

The most commonly used procedure for obtaining double-stranded cDNA bears two disadvantages. cDNA sequences corresponding to the 5'-end of the mRNA are lost, when S1 nuclease is used to digest the hairpin loop. This loop joins the two strands of DNA resulting from the self-priming of second strand synthesis. Furthermore, sequences have been observed in cloned cDNAs which do not occur in the mRNA sequence (7,33). As we reported earlier (7), the lysozyme cDNA clone pls-1 shows six nucleotides of unknown origin between the BamHI-linker and the cloned region of the lysozyme mRNA sequence. Richards et al. (33) reported similar observations for several cloned chicken ß-globin cDNAs. It was proposed (7,33) that these sequences, which are not present in the mRNA, were generated during the construction of blunt ended cDNA. The six aberrant nucleotides of pls-1 were very likely derived from its inverted and complementary counterpart at the very 5'-end of the lysozyme mRNA (Fig. 6)

To avoid loss of 5'-untranslated mRNA sequences in cDNA clones and artifacts such as described above, we decided to use a method to synthesize ds-cDNA (1,14) which does not involve S1 nuclease digestion. An oligo-homopolymer dCMP-tail was added to the 3'-end of full length cDNA by terminal transferase. Oligo$(dG)_{12-18}$, hybridized to the 3'-oligo(dC) tract, was used as a primer for the synthesis of the second DNA strand. The size of the cDNA was tested after the various reactions to ensure yields of full length cDNA (see Fig. 2). The lysozyme specific cDNA consisted of about 75 % full length molecules, whereas cDNAs derived from larger mRNAs such as ovomucoid or ovalbumin mRNA gave a lower yield of full length molecules.

To get high transformation efficiencies for the recombinant molecules, it was necessary to choose conditions in which the length of the homopolymer-tail did not exceed 10 nucleotides. cDNA and vector should bear tails of equal size (20). One ng of ds-cDNA gave rise to $5 \times 10^3$ recombinants in E.coli 5K. Although the strain is $recA^+$, we never observed instability

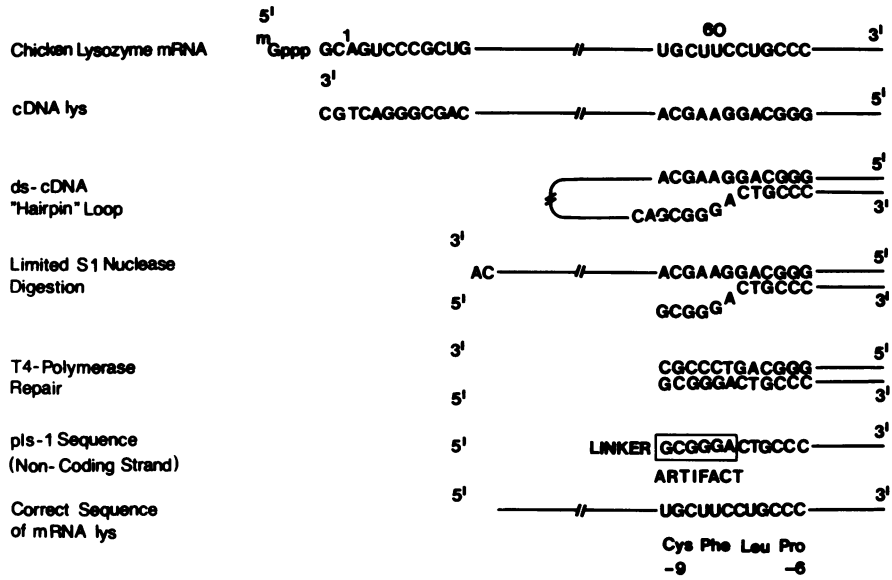


Fig. 6: Possible explanation for the loss of 5'-terminal sequences of eukaryotic mRNA during construction of ds-cDNA and concomitant introduction of artifacts (33).
As an example, the construction of the lysozyme-cDNA plasmid pls-1 (7) is shown. A lysozyme mRNA starting at position -2 of the genomic DNA (Table 2) is reverse transcribed into cDNA. A full length lysozyme cDNA molecule folds back and creates the hairpin structure priming the second DNA strand synthesis. Limited S1 nuclease digestion of the hairpin-loop results in a duplex molecule terminating in two non-pairing strands. The strand complementary to the 5'-terminus of the mRNA is then removed by the 3'-5' exonuclease activity of $T_4$ polymerase and consequently re-synthesized using the other strand as a template. This results in the loss of the sequences corresponding to the 5' terminal sequences of the mRNA as well as in the introduction of a six nucleotide artificial sequence in place of nucleotides coding for amino acids -9 and -8 of prelysozyme. The complementary and inverted counterpart of the artifact sequence of pls-1 appears at the very 5'-end of the lysozyme mRNA.

of the inserted cDNA. Starting from a non size-selected plasmid-library, 5 % of the clones were identified as carrying lysozyme specific cDNA. Sequence analysis showed that at least nine out of nineteen further characterized plasmids carry the entire 5'-end of the mRNA sequence (Fig. 4 and 5).

Multiple mRNAs with length heterogeneity of the 5'-untranslated region are generated from the lysozyme gene (M. Grez et al., submitted). They start at three different positions. Two of these positions are represented by nine

cloned cDNAs which have been sequenced (see Table 2). The clones can be divided into two groups according to the starting nucleotide. The third start-site is represented by a cDNA clone, pls-1124, carrying a sequence corresponding to the 5'-untranslated part of the mRNA which extends 22 or 24 nucleotides, respectively, further upstream (Table 2). Sequencing of the 5'-untranslated region of the lysozyme mRNA by reverse transcription under chain termination conditions revealed three mRNAs, the ends of which are equivalent to those found for cloned cDNA starts 1 to 3 (Fig. 5). It is unlikely that the different 5'-end sequences of lysozyme mRNA are due to premature termination of reverse transcriptase. AT-rich regions with similarities to the Goldberg-Hogness box are found in the genomic DNA 26 or 24 nucleotides upstream from all observed cDNA ends (M. Grez et al.,submitted). S1 nuclease mapping experiments with lysozyme mRNA, as well as nuclear RNA (M. Grez et al., submitted) and sequence analysis of capped oligonucleotides from isolated lysozyme mRNA (R.E. Rhoads, personal communication) are in excellent agreement with the three types of cDNA clones observed.

ACKNOWLEDGEMENTS

We wish to thank Dr. J.W. Beard for providing reverse transcriptase and Dr. W. Roewekamp for the gift of tailed pBR322, Drs. A.E. Sippel and E. Scherzinger for helpful discussions. We thank Drs. U. Bernard, W. Keller, G. Scherer and A. Scott, P. Shaw, C.M. Strange for critical comments on the manuscript as well as U. Joa for secretarial assistance. Part of this work was carried out at the Max-Planck-Institute für molekulare Genetik, Berlin.

+Present address: GBF, Abteilung Genetik, Mascheroder Weg 1, D-3300 Braunschweig, F.R.G.

REFERENCES

1 Rougeon, F., Kourilsky, P. and Mach, B. (1975) Nucl.Acids Res. 2, 2365-2378.
2 Rabbits, T.H. (1976) Nature 260, 221-225.
3 Wood, K.O. and Lee, J.C. (1976) Nucl.Acids Res. 3, 1961-1971.
4 Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976) Cell 8, 163-182.
5 Rougeon, F. and Mach, B. (1976) Proc.Natl.Acad.Sci. USA 73, 3418-3422.

6 Higuchi, R., Paddock, G.V., Wall, R. and Salser, W. (1976) Proc.Natl. Acad.Sci. USA 73, 3146-3150.
7 Sippel, A.E., Land, H., Lindenmaier, W., Nguyen-Huu, M.C., Wurtz, T., Timmis, K.N., Giesecke, K. and Schütz, G. (1978) Nucl.Acids Res. 5, 3275-3294.
8 Humphries, P., Chochet, M., Krust, A., Gerlinger, P., Kourilsky, P. and Chambon, P. (1977) Nucl.Acids Res. 4, 2389-2406.
9 Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J. and Goodman, H.M. (1977) Science 196, 1313-1319.
10 Seeburg, H.P., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) Nature 270, 495-499.
11 Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
12 Chang, A.Y., Ehrlich, H.A., Gunsalus, R.P., Nunberg, J.H., Kaufman, R.J., Schimke, R.T. and Cohen, S.N. (1980) Proc.Natl.Acad.Sci. USA 77, 1422-1446.
13 Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) Cell 16, 851-861.
14 Cooke, N.E., Coit, D., Weiner, R.J., Baxter, J.D. and Martial, J.A. (1980) J.Biol.Chem. 255, 6502-6510.
15 Hynes, N.E., Groner, B., Sippel, A.E., Nguyen-Huu, M.C. and Schütz, G. (1977) Cell 11, 923-932.
16 Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucl.Acids Res. 3, 863-877.
17 Gillespie, D. and Vogelstein, B. (1979) Proc.Natl.Acad.Sci. USA 76, 615-619.
18 Lindenmaier, W., Nguyen-Huu, M.C., Lurz, R., Blin, N., Stratman, M., Land, H., Jeep, S., Sippel, A.E. and Schütz, G. (1979) Nucl.Acids Res. 5, 1221-1232.
19 Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Proc.Natl.Acad.Sci. USA 74, 5350-5354.
20 Roewekamp, W. and Firtel, R.A. (1980) Develop.Biology 79, 409-418.
21 Hubatek, J. and Glover, S.W. (1970) J.Mol.Biol. 50, 111-127.
22 Dagert, M. and Ehrlich, S.D. (1979) Gene 6, 23-28.
23 Seif, J., Khoury, G. and Dhar, R. (1980) Nucl.Acids Res. 8, 225-2239.
24 Schreier, P.H. and Cortese, R. (1979) J.Mol. Biol. 129, 169-172.
25 Jung, A., Sippel, A.E., Grez, M. and Schütz, G. (1980) Proc.Natl.Acad. Sci. USA 77, 5759-5763.
26 Stein, J.P., Catterall, J.F., Woo, S.L.C., Means, A.R. and O'Malley, B.W. (1978) Biochemistry 17, 5763-5772.
27 McReynolds, L.A., Catterall. J.F. and O'Malley, B.W. (1977) Gene 2, 217-231.
28 Sidikaro, J. and Nomura, M. (1975) J.Biol.Chem. 250, 1123-1131.
29 Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475.
30 Tsujimoto, Y. and Suzuki, Y. (1979) Cell 16, 425-436.
31 Tsujimoto, Y. and Suzuki, Y. (1979) Cell 18, 591-600.
32 Weaver, R.F. and Weissmann, C. (1979) Nucl.Acids Res. 7, 1175-1193.
33 Richards, R.J., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) Nucl.Acids Res. 5, 1137-1146.