Volume 9 Number 7 1981 Nucleic Acids Research

Heterogeneity of the 5' terminus of hen ovalbumin messenger ribonucleic acid

Lawrence T.Malek*, William H.Eschenfeldt*, Theodore W.Munns+ and Robert E.Rhoads*

*Department of Biochemistry, University of Kentucky, Lexington, KY 40536, and +Washington University School of Medicine, Washington University, St. Louis, MO 63110, USA

Received 1 December 1980

ABSTRACT

The 5'-terminal sequence of hen ovalbumin mRNA was investigated using a novel labeling method. Ovalbumin mRNA was purified by hybridization to complementary DNA coupled to cellulose. The mRNA thus purified was shown to be 97.9% pure by hybridization with plasmid DNA containing sequences to the messengers coding for conalbumin and ovomucoid, the next two most abundant messengers of oviduct. After digestion with RNase T_1 and alkaline phosphatase, 5'-terminal capped oligonucleotides were selected by binding to anti-m^7G-Sepharose. These were then labeled using RNA ligase and [5'-^{32}P]pCp, separated by two-dimensional gel electrophoresis, and sequenced by partial digestion with base-specific ribonucleases. A nested set of three capped oligonucleotides was identified. Their structures and relative abundances were m^7GpppAUACAG, 3%; m^7GpppACAUACAG, 61%; and m^7GpppGUACAUACAG, 36%.

INTRODUCTION

In an earlier study (1), we observed that the messenger coding for ovalbumin ($mRNA_{OV}$) was terminated by two kinds of caps: m^7GpppA, representing approximately 2/3 of the total, and m^7GpppG, representing 1/3. (All of the typical methylated derivatives were present as well.) The A-containing caps were predicted from the 5'-terminal sequence of $mRNA_{OV}$, determined by McReynolds *et al.* (2) by primer extension in the presence of dideoxynucleoside triphosphates. However, the origin of the G-containing caps remained obscure. Possible origins included contamination of the $mRNA_{OV}$ preparation with other mRNAs, allelism in the ovalbumin gene (3), and heterogeneous initiation sites for transcription of $mRNA_{OV}$. To distinguish between these possibilities, we set out to sequence the oligonucleotides to which each type of cap was attached.

Direct sequencing of an unlabeled RNA, such as $mRNA_{OV}$, requires a method for the introduction of radioactivity. However, our previous attempts to label the 5'-terminus of $mRNA_{OV}$ with polynucleotide kinase and [γ-^{32}P]ATP were unsuccessful (4). We therefore developed a new method for labeling 5'-termi-

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

nal oligonucleotides of mRNA which uses anti-m^7G-Sepharose to selectively purify m^7G-containing oligonucleotides, and RNA ligase and [5'-^{32}P]cytidine-3', 5'-bisphosphate (pCp) to introduce the label. This method should be generally applicable for the determination of 5'-terminal structure in mRNA, and should complement other methods currently in use. Applied to $mRNA_{ov}$, it revealed the presence of a variety of 5'-terminal oligonucleotides. The three most abundant of these were of different lengths but had the same sequence, reading from the 3' terminus. This suggests different processing sites of a larger precursor, or heterogeneity of transcription initiation sites for $mRNA_{ov}$.

METHODS

Purification of Plasmids and Inserts

E. coli K12 C600 strains, containing the ovalbumin plasmid pCRlov2.1 and the conalbumin plasmid pCON-1, were generously supplied by Dr. Pierre Chambon, Strasbourg, France. The χ1776 strain, containing the ovomucoid plasmid pOM-48 was a gift from Dr. Günther Shütz, Heidelburg, GFR. Growth of cells and isolation of plasmids (5) were performed in compliance with NIH guidelines. Digestions with restriction endonucleases *Bam HI*, *Hind III*, *Hha I* and *Msp I* (New England Biolabs) were each performed using the supplier's assay conditions. A 2.4 kb fragment containing the ovalbumin insert of pCRlov2.1 was purified by digestion with *Msp I* and *Hha I* followed by sucrose gradient centrifugation. Similarly, the 0.41 kb ovomucoid insert was purified from a *Bam HI* digest of pOM-48. The plasmid pCON-1, linearized by digestion with *Hind III*, together with the ovalbumin and ovomucoid inserts were labeled by nick-translation (6) with either [^{3}H]- or [^{32}P]-deoxynucleoside triphosphates (Amersham).

Preparation of DNA-Cellulose

Diazobenzyloxymethyl (DBM)-cellulose was prepared as described by Noyes and Stark (7). The ovalbumin DNA insert was coupled to DBM-cellulose using the procedure of Noyes and Stark (7), with the modifications of Childs *et al.* (8). Thirty to sixty micrograms of DNA in 80% dimethyl sulfoxide and 5 mM sodium phosphate, pH 6.5, were coupled to 10 mg DBM-cellulose for 2 d at 4°C. The DNA-cellulose was washed to remove uncoupled DNA (7), and pretreated to reduce non-specific binding of RNA (9). Coupling efficiency of DNA to cellulose, typically 50 to 80%, was determined by including a small amount of ^{32}P-labeled nick-translated DNA.

Purification and Assay of $mRNA_{ov}$

Ovalbumin mRNA was obtained by hybridization to DNA-cellulose using the procedure of Childs *et al.* (8) with a few modifications. Forty to fifty micrograms of hen oviduct mRNA (1) in 0.2 ml of RNA hybridization buffer (50% formamide, 10 mM Tris-HCl, pH 7.4, 0.6 M NaCl, 1 mM EDTA, 0.5 mg/ml poly(A)) was hybridized to 10 mg DNA-cellulose for 10 min at 52°C. The cellulose was then washed with RNA hybridization buffer and 2x SSC (8). The mRNA was eluted with 100% formamide at 37°C and precipitated twice with ethanol. A trace amount of oviduct [^{3}H]mRNA, prepared as described by Malek *et al.* (1), was included in the hybridization mixture to determine the yield of $mRNA_{ov}$ for each hybridization and to monitor subsequent purification of m^7G-containing oligonucleotides. Typically, 10 to 15 µg of purified $mRNA_{ov}$ were obtained from each hybridization.

The purity of $mRNA_{ov}$ was determined by hybridization of ^{3}H-labeled nick-translated plasmid DNA probes to RNA which had been coupled to filters. Nitrobenzyloxymethyl-paper was prepared as described by Alwine *et al.* (10) and stored at 4°C. Immediately before use, the paper was cut into 1 cm diameter circles and converted to DBM-paper. A 5 µl-solution containing 25 mM sodium phosphate, pH 6.5, 2 mg/ml yeast RNA, and varying amounts (from 1 to 500 ng) of either oviduct mRNA or purified $mRNA_{ov}$, was spotted on each filter. After incubation overnight at 4°C, the filters were pretreated by incubation in two changes of DNA hybridization buffer (50% formamide, 0.9M NaCl, 50 mM sodium phosphate, pH 7.0, 5 mM EDTA, 200 µg/ml bovine serum albumin, 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone) plus 0.1% SDS and 1.0% glycine at 42°C for two hours each. As many as 35 filters were then hybridized to ^{3}H-labeled plasmid DNA in a plastic boiling bag containing 0.5 ml of DNA hybridization buffer for 18 h at 42°C. After hybridization all filters were washed twice in 0.36 M NaCl, 20 mM sodium phosphate, pH 7.0, 2 mM EDTA, 0.1% SDS at room temperature for 15 minutes followed by two washes in 18 mM NaCl, 1 mM sodium phosphate, pH 7.0, 0.1 mM EDTA, 0.1% SDS at 52°C for 20 minutes. After treating the filters for 1 h at 60°C with 0.2 ml of NCS (Amersham) per filter, 4 ml of toluene-based scintillation fluid were added and radioactivity derermined by liquid scintillation spectrometry.

Preparation of m^7G-Containing Oligonucleotides

Ovalbumin mRNA (10 to 15 µg) was digested to completion in a 20 µl reaction containing 50 mM Tris-HCl, pH 8.0, and 50 µg/ml RNase T_1 (Calbiochem). After incubating the reaction mixture for 1 h at 37°C, 0.06 units of calf intestine

alkaline phosphatase (Boeringer) were added and incubation was continued for 30 minutes at 37°C. The reaction mixture was extracted twice with phenol: chloroform (1:1), twice with chloroform, and once with ether. The oligonucleotides were evaporated to dryness and resuspended in 400 µl PBS (0.01 M sodium phosphate pH 7.0, 0.15 M NaCl). The oligonucleotides were added to 25 µl (wet volume) of anti-m^7G-Sepharose, prepared as described by Munns *et al.* (11), and allowed to bind by maintaining the Sepharose in suspension for 40 min at room temperature. The suspension was centrifuged at 1000 x g for 10 s. After removing the supernatant and washing the Sepharose with five 400 µl-volumes of PBS, the bound oligonucleotides were eluted by incubation in 400 µl of a solution containing 7 M urea, 1 mg/ml 7-methylguanosine, and 0.1 M sodium acetate pH 5.0, for 30 min at 37°C. The eluted oligonucleotides were desalted on a column (1.0 x 20 cm) of Sephadex G-10 and evaporated to dryness (1).

^{32}P-Labeling of Oligonucleotides

The m^7G-containing oligonucleotides were labeled using T_4 RNA ligase as described by Bruce and Uhlenbeck (12). Briefly, the desalted oligonucleotides (2 to 5 pmol) and 25 µCi [5'-^{32}P]pCp (Amersham, 2-3000 Ci/mmol) were added to a 20 µl-reaction of 50 mM HEPES pH 7.5, 20 mM $MgCl_2$, 3 µM ATP 3.3 mM dithiothreitol, 10 µg/ml bovine serum albumin, 10% dimethylsulfoxide, and 200 units/ml RNA ligase (P. L. Biochemicals). The reaction was incubated for 16 h at 5°C. The ^{32}P-labeled oligonucleotides were purified by binding to anti-m^7G-Sepharose, and the eluted material was made to 100 µg/ml in yeast RNA and precipitated with ethanol. For analysis of the nucleoside residue nearest to the added pCp, 1 µl of the RNA ligase reaction mixture or 0.5 µl of the purified [^{32}P]oligonucleotides were digested with RNase T_2 (1) and analyzed by high performance liquid chromatography (HPLC) on a Whatman Partisil SAX column (13).

Two-Dimensional Electrophoresis

Two-dimensional gel electrophoresis was performed as described by DeWachter and Fiers (14), as modified by Lockard *et al.* (15). The ethanol-precipitated [^{32}P]oligonucleotides were dissolved in 10 µl H_2O, and to this was added 5 µl of a layering solution containing 50% sucrose, 10 M urea, 0.2% xylene cyanol FF and 0.2% bromophenol blue. Electrophoresis in the first dimension was performed using a gel (40 x 15 x 0.15 cm) of 10% acrylamide, 0.5% N,N'-methylene bisacrylamide, 7 M urea and 25 mM citric acid pH 3.5. The gel was polymerized (14) and pre-electrophoresed at 1000 V for 1 h.

After loading the sample, electrophoresis was continued at 1000 V and 5°C for about 6 h, until the bromophenol blue dye migrated 19 cm. A gel strip containing all of the radioactivity was removed and placed near the bottom of a mold (40 x 30 x 0.15 cm). The gel for the second dimension, containing 20% acrylamide, 1% N,N'-methylenebisacrylamide, 7 M urea, 90 mM Tris-borate pH 8.3, 2.5 mM EDTA, 0.1% ammonium persulfate and 0.2% N,N,N',N'-tetramethylethylenediamine (TEMED) was poured into the mold and allowed to polymerize. Electrophoresis in the second dimension was performed at 600 V and room temperature until the bromophenol blue dye migrated 26 cm. An autoradiogram, using Kodak No-screen NS-5T X-ray film, was used to locate the [^{32}P] oligonucleotides, which were eluted and precipitated by the method of DeWachter and Fiers (14).

RNA Sequencing

[^{32}P]oligonucleotides (10^4 cpm), containing 5 to 25 µg carrier RNA, were added to 40 µl of 20 mM sodium acetate pH 5.0, and 100 µg/ml carrier RNA. After removing a 3 µl-aliquot, 0.1 µg of RNase T_1 was added to the remainder. From this, 3 µl were removed and mixed with 0.5 units tobacco acid pyrophosphatase (16). The T_1 digest, along with the two aliquots described above, were incubated at 37°C for 30 min. To three 5 µl-aliquots of the T_1 digest were added 0.5 µl of either RNase A (Sigma, 3 x 10^{-4} µg/ml), RNase Phy I (Enzo Biochem, 10^3 units/ml), or RNase U_2 (Calbiochem, 5 units/ml). A 0.5 µl-aliquot was then removed from each of these and added to a fresh 5 µl-aliquot of the T_1 digest. The partial RNase digests were incubated at 20°C for 10 min, quenched by addition of 2 µl of the layering solution described above, and maintained at -70°C until electrophoresis. Electrophoresis was performed on gels (40 x 15 x 0.04 cm) composed of 25% acrylamide, 0.83% N,N'-methylenebisacrylamide and 8 M urea, as described by Sanger and Coulson (17). Electrophoresis proceeded at 800 V until the xylene cyanol FF and bromophenol blue dyes migrated to 10 and 19 cm, respectively. The gel was subjected to autoradiography on X-ray film, using either Kodak No-screen film NS-5T or pre-flashed (18) Kodak X-Omat XR5 film with Dupont Cronex Hi-plus intensifying screens.

Analysis of [^{3}H]mRNA$_{OV}$

[^{3}H]mRNA$_{OV}$ (4.7 x 10^4 cpm) was digested in a 20 µl-reaction mixture containing 20 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mg/ml oviduct polysomal RNA and 100 µg/ml RNase T_1. The reaction mixture was incubated 1.5 h at 37°C and then extracted with phenol:chloroform and ether. After adding 10 µl of layer-

ing solution, the [^{3}H]oligonucleotides were loaded onto a gel (40 x 15 x 0.15 cm) containing 25% acrylamide, 0.83% N,N'-methylenebisacrylamide, 8 M urea, 90 mM Tris-borate pH 8.3, 2.5 mM EDTA, 0.1% ammonium persulfate and 0.2% TEMED. Electrophoresis was performed at 400 V until the bromophenol blue dye migrated 15 cm. A strip (1.0 x 25 cm), containing the radioactivity was removed and sliced into 2.2 mm pieces. After treating with 0.5 ml of 90% NCS overnight at 37°C, 4 ml of toluene-based scintillation fluid were added and radioactivity was determined as above.

RESULTS

Purity of mRNA$_{ov}$

To exclude the possibility that the G-containing caps were contributed by contaminating mRNAs, mRNA$_{ov}$ was purified by hybridization of hen oviduct mRNA to the complementary insert of the plasmid pCRlov2.1 coupled to cellulose (DNA-cellulose). The purity of mRNA$_{ov}$ was determined by first immobilizing the RNA on filters and then hybridizing to ^{3}H-labeled plasmid probes for mRNA$_{ov}$ and the next two most abundant oviduct mRNAs, those coding for conalbumin (mRNA$_{con}$) and ovomucoid (mRNA$_{om}$). By comparing the hybridization of the probes to purified mRNA with the hybridization of the probes to a known standard, oviduct mRNA, the concentrations of each component of the purified mRNA mixture were calculated. The relative abundance of each mRNA component was proportional to the slopes of the hybridization curves shown in Fig. 1. The relative abundance of mRNA$_{ov}$ in the purified mRNA (Fig. 1A) was nearly twice that found in the oviduct mRNA starting material (solid line). Conversely, levels of mRNA$_{con}$ (Fig.1B) and mRNA$_{om}$ (Fig.1C) were both diminished about 24-fold by purification on DNA-cellulose. The overall purity of mRNA$_{ov}$ was 97.9%, based on calculations outlined in Table 1. The most abundant contaminant, conalbumin mRNA, was only 0.4% of the RNA.

Heterogeneity of T$_1$ Oligonucleotides of [^{3}H]mRNA$_{ov}$

Previously, we observed several capped RNase T$_1$-derived oligonucleotides in partially purified mRNA$_{ov}$ that had been labeled with tritium in the 5'-terminal m^{7}G and 3'-terminal A residues (1). To determine whether the heterogeneity persisted in a more purified preparation, [^{3}H]mRNA$_{ov}$ obtained by hybridization to DNA-cellulose was digested with RNase T$_1$ and analyzed by polyacrylamide gel electrophoresis (Fig. 2). Most of the label was in the T$_1$ fragment derived from the poly(A) tract, presumably due to cleavage of the mRNA during purification or labeling, followed by selection of only 3'-

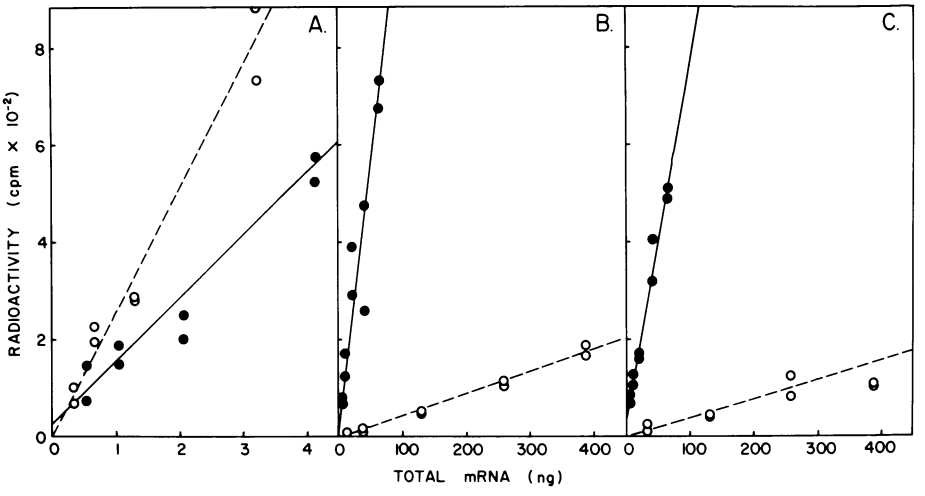

Fig. 1: Hybridization of Oviduct and Purified mRNA to ^{3}H-labeled Plasmid Probes. A) Increasing amounts (1 to 10 ng) of hen oviduct (o—) or purified (o—) mRNA were immobilized on filters and hybirdized to 2.1 x 10^5 cpm of ^{3}H-labeled pCR1ov2.1 insert (1.1 x 10^7 cpm/µg), as described in "Methods". The concentration of oviduct mRNA was measured assuming 25 A_{260}/mg and that of purified mRNA was calculated as described in Table 1. B) Same as Fig. 1A, except 2.2 x 10^5 cpm of ^{3}H-labeled pCON-1 (5.0 x 10^6 cpm/µg) were used as a probe and 1 to 500 ng purified mRNA were immobilized on the filters. C) Same as Fig. 1B, except 9.6 x 10^4 cpm of ^{3}H-labeled pOM-48 insert (3.1 x 10^6 cpm/µg) were used as a probe.

termini by oligo(dT)-cellulose. However, 15% was in four or more small, discrete oligonucleotides. The length of each was determined by comparison with the 5'-terminal RNase T_1 oligonucleotides of α- and β-globin mRNA, which have structures of $m^7Gppp(Np)_9$ and $m^7Gppp(Np)_7$, respectively. The major oligonucleotide, contributing 61% of the total label in m^7G, had the structure $m^7Gppp(Np)_8$ and was called Ov 8. The remaining radioactivity in m^7G was contained in three oligonucleotides having the structures $m^7Gppp(Np)_{10}$ (Ov 10), $m^7Gppp(Np)_6$ (Ov 6), and $m^7GpppNp$ (Ov 1), with relative amounts of 25%, 3%, and 11%, respectively. The relative abundance of oligonucleotides obtained was consistent through several experiments, as demonstrated by acrylamide gel analysis as well as by DEAE Sephacel chromatography (1). We previously identified the cap structures of Ov 10 as $m^7GpppGm$, and that of Ov 8 as $m^7Gppp(m\ )A(m)$, where parentheses indicate partial methylation (1, 21). Ov 1 had the structure $m^7GpppGp$, which could have originated from an

TABLE 1. Calculation of $mRNA_{ov}$ Purity

mRNA Assayed	Hybridization Probe	Oviduct mRNA		Purified mRNA		
		Specific Hybridization	Relative Abundance	Specific Hybridization	Calculated RNA Concentration	Relative Abundance[c]
		cpm/ng	%	*cpm/μl*	*μg/ml*	%
Ovalbumin	pCR1ov2.1	258	50[a]	63,200	245	97.9
Conalbumin	pCON-1	106	10[b]	113	1.07	0.43
Ovomucoid	pOM-48	115	6.6[a]	77	0.67	0.27

[a]Hynes et al. (19).

[b]Lee et al. (20).

[c]Calculated by assuming that the ratios of $mRNA_{con}$ and $mRNA_{om}$ to other non-ovalbumin mRNAs remain constant during purification, i.e., that $mRNA_{con}$ and $mRNA_{om}$ behave as typical contaminants. Accordingly, if $mRNA_{con}$ represents 20% of the non-ovalbumin mRNA in oviduct mRNA, the purified sample contains 1.07/0.2 = 5.35 μg/ml of non-ovalbumin mRNA (2.1%). Similarly, if $mRNA_{om}$ represents 13.2%, then the level of non-ovalbumin mRNA is 0.67/0.132 = 5.08 μg/ml (2.0%). The relative abundance of $mRNA_{ov}$ is calculated from the average of these two values.

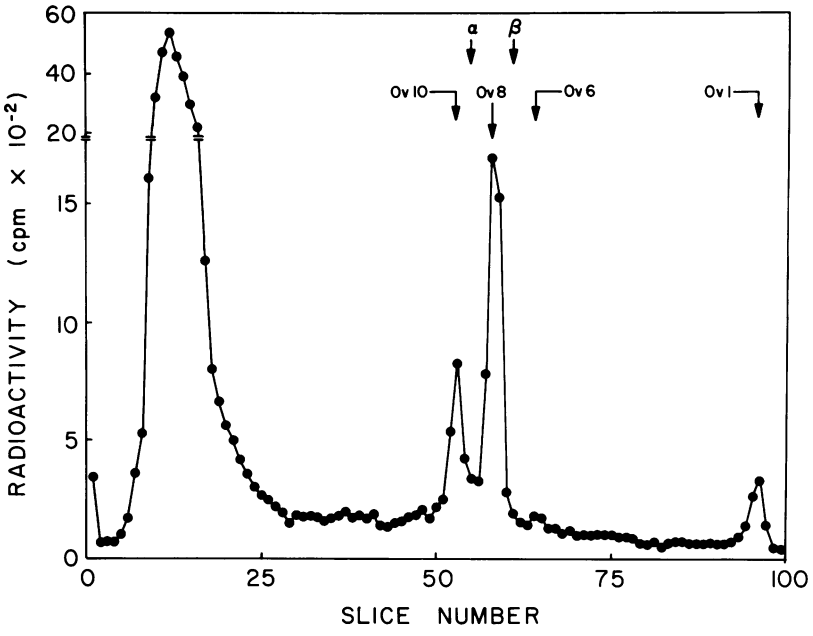

Fig. 2: Polyacrylamide Gel Electrophoresis of [^{3}H]mRNA$_{ov}$ Digested with RNase T_1. [^{3}H]mRNA$_{ov}$ (10^5 cpm), purified from oviduct [^{3}H]mRNA (2.4 x 10^4 cpm/µg) on DNA-cellulose, was digested with RNase T_1 and separated by electrophoresis on 20% polyacrylamide gels containing 7 M urea as described in "Methods". The gel was sliced and counted to determine radioactivity. The positions of the four ovalbumin mRNA 5'-T_1 oligonucleotides are indicated by arrows. The migration of the 5'-terminal oligonucleotides of α and β globin mRNA, run on a parallel lane, is also indicated.

oligonucleotide with the same structure as Ov 10, but without 2'-O-methylation. If Ov 10 and Ov 1 are derived from the same mRNA, their sum would represent 36% of the total.

Labeling of T_1 Oligonucleotides

In preparation for sequence analysis, the 5'-terminal T_1 oligonucleotides were end-labeled using RNA ligase and [5'-^{32}P]pCp. Acceptor oligonucleotides for RNA ligase were prepared by digesting mRNA$_{ov}$ with RNase T_1 and alkaline phosphatase. To conserve label, the 5'-terminal oligonucleotides were first purified from the internal ones with anti-m^{7}G-Sepharose. The conditions used for binding and elution of oligonucleotides gave near quantitative recovery of m^{7}G, based on the recovery of 12 to 15% of the tritium label in [^{3}H]mRNA$_{ov}$. The 5'-terminal oligonucleotides were then labeled with [5'-^{32}P]pCp and pur-

ified again on anti-m^7G-Sepharose. About 20% of the label was incorporated into oligonucleotides, most of which was bound to anti-m^7G-Sepharose the second time. To verify that the oligonucleotides were authentic T_1-digestion products and not derived by random hydrolysis, an aliquot of the bound [^{32}P]oligonucleotides was further digested with RNase T_2 and analyzed by HPLC. Most of the radioactivity (92%) was found in Gp, indicating that some random hydrolysis had occurred but the T_1-oligonucleotides were predominantly intact.

Separation of 5'Terminal Oligonucleotides

End-labeled, m^7G-containing T_1 oligonucleotides of $mRNA_{OV}$ were separated by two-dimensional polyacrylamide gel electrophoresis (Fig. 3A). Electrophoresis in the first dimension (from left to right) was on a 10% polyacrylamide gel at pH 3.5. Separation is primarily by base composition in this dimension, with oligonucleotides rich in G and U migrating faster than those with A and C residues. Electrophoresis in the second dimension (from bottom to top) was on a 20% polyacrylamide gel at pH 8.3, where separation is by size. The sizes of the most prominent oligonucleotides were determined by comparison with the positions of α- and β-globin mRNA 5'-terminal T_1-oligonucleotides, labeled by the same method and analyzed on a similar two-dimensional gel (Fig. 3B). The designations Ov 6, Ov 8, and Ov 10 were given to the oligonucleotides that correspond in length to the respective [^3H]oligonucleotides of Fig. 2. The relative intensities of spots varied among experiments, (40-70% Ov 8, 2-4% Ov 6) indicating possible degradation occurring at some stage of the labeling procedure. The intensity of Ov 10 was especially variable, ranging from 1 to 6% of the total radioactivity. In the experiment represented in Fig. 3A, it was present in amounts less than expected from Fig. 2. Also, spots corresponding to oligonucleotides which were lacking either a cap or a 3'-terminal G residue were occasionally found. The actual number and intensities of these spots were variable. In this experiment, a total of 12 spots were sequenced. However, only those spots corresponding to Ov 6, Ov 8, and Ov 10 contained both a cap and a 3'-terminal G. By sequence analysis (data not shown) the other spots, such as that indicated by Ov 4', appeared to be degradation products of the 3 parent oligonucleotides.

The complexity of the oligonucleotide pattern observed with purified $mRNA_{OV}$ (Fig.3A) raises the possibility of contamination by oligonucleotides of other oviduct mRNAs. For comparison, a sample of oviduct mRNA was subjected to the same procedure (Fig.3C). It is clear that hybridization to ovalbumin DNA-cellulose resulted in the selection of a discrete set of oligonucleotides.

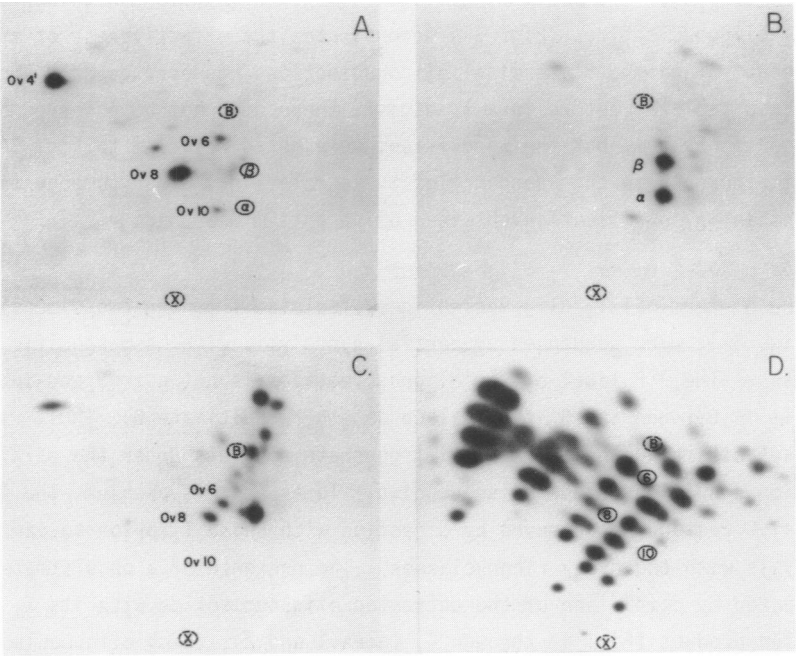

Fig. 3: Two-Dimensional Gel Electrophoresis of RNase T_1 [^{32}P]oligonucleotides. A) 5'-Terminal oligonucleotides were purified from an RNase T_1 and alkaline phosphatase digest of $mRNA_{ov}$ (15 µg) using anti-m^7G-Sepharose and labeled using RNA ligase and [^{32}P]pCp, as described in "Methods". After purification again on anti-m^7G-Sepharose, the [^{32}P]oligonucleotides (2 x 10^6 cpm) were separated by two-dimensional gel electrophoresis as described in "Methods". Migrations of xylene cyanol FF (X) and bromophenol blue (B), as indicated, were 17 and 19 cm, respectively, for the first dimension, and 14 and 26 cm, respectively, for the second. The migrations of the 5'-terminal T_1 oligonucleotides of α- and β-globin mRNA are indicated as α and β (see Fig. 3A). Only the top portion of the gel, containing all the radioactivity is shown. B) 5'-Terminal T_1 oligonucleotides were purified from a digest of rabbit globin mRNA (5 µg) and labeled as described above. The [^{32}P]oligonucleotides (1 x 10^6 cpm) were separated as above. C) 5'-Terminal T_1 oligonucleotides were purified from a digest of oviduct mRNA (75 µg) and labeled, as above. The [^{32}P]oligonucleotides (5 x 10^6 cpm) were separated as above. D) T_1 oligonucleotides from a digest of $mRNA_{ov}$ (1 µg) were labeled as above, but without purification on anti-m^7G-Sepharose. The [^{32}P]oligonucleotides (6 x 10^6 cpm) were separated as above.

Interestingly, although $mRNA_{ov}$ is the major component of oviduct mRNA (50%), its oligonucleotides are not the predominant ones in oviduct mRNA. A second hybridization of $mRNA_{ov}$ to DNA-cellulose failed to reduce the complexity of oligonucleotides observed in Fig. 3A (data not shown).

Comparison of m^7G-containing oligonucleotides (Fig. 3A) with unfractionated T_1 oligonucleotides (Fig. 3C) demonstrates the effectiveness of anti-m^7G-Sepharose. Since the 5'-terminal oligonucleotides represented only about 0.2% of the available acceptors in a T_1 digest, they could not be observed on the two dimensional gel, but their positions were noted relative to Fig. 3A. The dramatic increase in the abundance of these relatively minor components shows the substantial purification obtained using anti-m^7G-Sepharose.

Sequence of 5'-Terminal T_1 Oligonucleotides

The 5'-terminal T_1 oligonucleotides, isolated from the two-dimensional gel (Fig. 3A), were sequenced by partial hydrolysis with base-specific ribonucleases. The ^{32}P label of the oligonucleotides is not a true end-label since it is between the 3'-terminal Cp and the penultimate G. Thus, ribonucleases attacking the GpC bond will generate doublets under the partial digestion conditions used for sequencing. To avoid this problem, the 3'-terminal C residue was removed by digestion with RNase T_1 prior to partial hydrolysis with the other ribonucleases. The presence of a penultimate G was checked by comparison of the untreated oligonucleotide with its T_1 digestion product (Fig. 4A through C, lanes 1 and 2). Each oligonucleotide was shortened one nucleotide by treatment with RNase T_1, as expected for authentic T_1 oligonucleotides.

Degradation of 5'-terminal oligonucleotides would yield shorter oligonucleotides with a common sequence from the 3' terminus. To verify that these were authentic 5'-terminal oligonucleotides, the presence of a cap was demonstrated. An aliquot of the T_1-digest of each oligonucleotide was treated with tobacco acid pyrophosphatase (Fig. 4A through C, lane 3). All three oligonucleotides underwent a dramatic increase in mobility upon treatment with the pyrophosphatase, suggesting the removal of the 5'-terminal ppm^7G. Other oligonucleotides present in the two-dimensional gel (Fig. 3A) were lacking either a 5'-terminal cap or a 3'-penultimate G. These were probably degradation products since they had sequences common with the major three oligonucleotides. The most prominent of these, spot Ov 4', had the structure ACAG.

The sequences for Ov 6, Ov 8, and Ov 10 were deduced from the partial ribonuclease digestion patterns in lanes 4 through 9 of Figs. 4A through C. The ribonucleases used were RNase A, specific for CpN and UpN bonds, RNase Phy I, for bonds other than CpN, and RNase U_2, specific for ApN bonds. The sequences are the same from the 3' ends of each oligonucleotide (reading

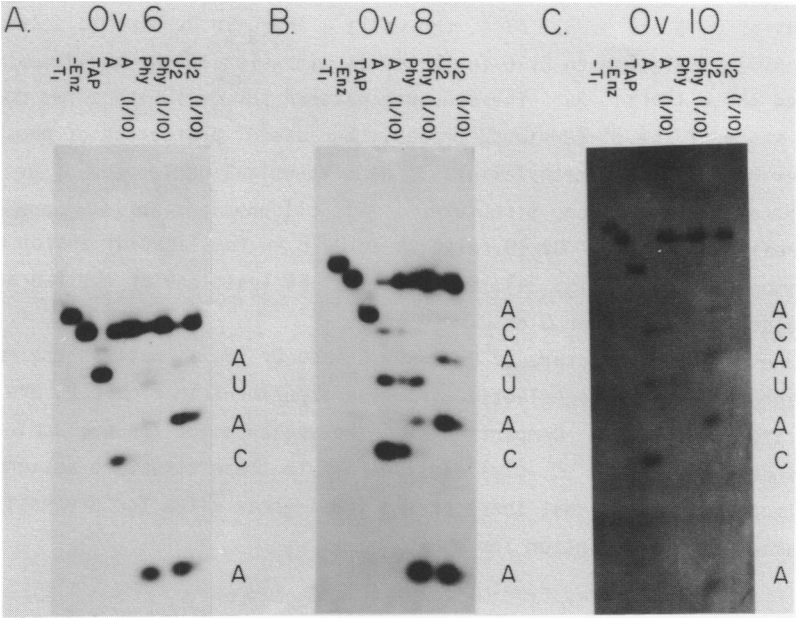

Fig. 4: Sequence Analysis of 5'-Terminal T_1 Oligonucleotides of $mRNA_{Ov}$. A) The Ov 6 oligonucleotide was eluted from the two-dimensional gel (Fig. 3A) and aliquots (1×10^4 cpm) were treated with various RNases to obtain its sequence as described in "Methods". The appropriate RNase digestion is indicated at the top of each lane. All samples were digested to completion with RNase T_1 except the one in the $-T_1$ lane. The samples in the other lanes were digested with no additional enzyme (-Enz), tobacco acid pyrophosphatase (TAP), RNase A (A, 1 and 1/10 dilutions), RNase Phy I (Phy, 1 and 1/10 dilutions), and RNase U_2 (U_2, 1 and 1/10 dilutions). The top 1/3 of the gels were not shown. Autoradiography was for 10 d using No-screen film. B) Same as Fig. 4A, except 1×10^4 cpm of Ov 8 oligonucleotide were used, and autoradiography was for 5 d. C) Same as Fig. 4A, except 2×10^3 cpm of Ov 10 were used. Autoradiography was for 10 d using "preflashed" X-Omat XR5 film with intensifying screens.

from bottom to top), but are extended to different lengths at the 5' ends. The sequences for Ov 6 and Ov 8 terminate at the cap as indicated by comparison of the largest partially digested fragments with the pyrophosphatase treated oligonucleotide (lane 3). We have previously shown that 14% of the caps in Ov 8 exist as cap 0 structures (1). Consequently the band representing a cleavage of the 5'-terminal A resulting from these structures (since forms containing 2'-O-methylation would render an adjacent phosphodiester bond resistant to RNase), is correspondingly lighter in intensity than the other

bands. The number of nucleotides in each sequence agrees with the predicted lengths of Ov 6 and Ov 8. Also, the extra A and C in Ov 8 would decrease its mobility relative to Ov 6 in the first dimension of the two-dimensional gel, as shown in Fig. 3A. The sequence pattern for Ov 10 stops two nucleotides short of the 5' terminus, probably because of protection of phosphodiester bonds by 2'-O-methylations. The 5'-terminal nucleotide of Ov 10 must be a G since the cap structure of this oligonucleotide is $m^7GpppGm$. The greater mobility of Ov 10 relative to Ov 8 in the first dimension of the two dimensional gel (Fig. 3A), suggests that at least one of the two additional nucleotides is either G or U.

The proposed structure of Ov 6, Ov 8, and Ov 10, as well as the relative abundance of each oligonucleotide from the experiment in Figure 2, are summarized in Table 2. Comparison with the sequence for the ovalbumin gene published by Gannon *et al.* (22) adds support to these structure assignments and strongly suggests that there are at least three sites for processing or initiation of transcription for $mRNA_{ov}$.

DISCUSSION

The sequence of Ov 8, the major 5'-terminal T_1 oligonucleotide, agrees with the sequences of $mRNA_{ov}$ (2, 23) and the ovalbumin gene (22). Indications of 5'-terminal heterogeneity for $mRNA_{ov}$ were observed previously. McReynolds *et al.* (2) found that when $mRNA_{ov}$, hybridized to a restriction fragment primer, was copied with reverse transcriptase, either in the presence or

TABLE 2. 5'-Terminal Structures of $mRNA_{ov}$

RNase T_1 Oligonucleotide	Structure	Abundance
Ov 6	$m^7GpppAUACAG$	3%
OV 8	$m^7GpppACAUACAG$	61%
OV 10	$m^7GpppGUACAUACAG$[a]	36%[b]
	···GTACATACAG···[c]	

[a]Assignment of U in third position from 5'-terminus is tentative and based primarily on the sequence of the gene.

[b]Includes 11% from m^7GpppG (Ov 1), assumed to arise from an Ov 10 form with no methylations.

[c]Sequence of the noncoding strand of the ovalbumin gene in the region of the cap site (22).

absence of dideoxynucleoside triphosphates, synthesis would always extend 1 to 2 nucleotides beyond the "full stop", which corresponded to the 5' terminus of the Ov 8 form of the mRNA. Several possible explanations, including 5'-heterogeneity as well as anomalous synthesis near the cap, were offered as a possible explanation for the apparent heterogeneity. Wasylyk *et al.* (24) digested a 5' end-labeled gene fragment, hybridized to $mRNA_{ov}$, with S_1 nuclease in an experiment that mapped the 5' terminus of the mRNA on the ovalbumin gene. Two S_1-resistant fragments were obtained. One terminated at the 5' terminus of the Ov 8 form, and the other, having only about half the label of the first, was two nucleotides longer. The second fragment agrees with the size and intensity of a fragment that would be protected by the Ov 10 form of the mRNA. However, a definite conclusion could not be reached since the S_1 mapping technique will occasionally give superfluous bands caused by incomplete S_1 digestion (24).

The sequencing method described in this paper complements other sequencing approaches, in that it avoids ambiguities near the cap. It also has advantages over the more conventional approach of removing the cap and labeling the 5' terminus with polynucleotide kinase and [γ-^{32}P]ATP. Not all mRNAs, including $mRNA_{ov}$, can be labeled with polynucleotide kinase (4). Also, the efficiency of labeling with polynucleotide kinase depends on the identity of the terminal nucleotide, 2'-O-methylations, and secondary structure. These problems are avoided with the RNA ligase method since the nucleotide labeled is generally distant from the cap. The most important advantage of this method is that the cap remains intact throughout the procedure, allowing verification of a capped oligonucleotide. The polynucleotide kinase method of 5' end-labeling will not discriminate between 5' termini which are the result of random degradation, and those which result from decapping of authentic 5' termini. Using the polynucleotide kinase labeling method, it would have been impossible to conclude that Ov 6 was a unique 5' terminus for $mRNA_{ov}$ since it could have been a degradation product of Ov 8.

It should be noted that differences were observed in the quantitation of oligonucleotides labeled with [5'-^{32}P]pCp and RNA ligase when compared with the same sample labeled with NaB[^{3}H]$_4$. Variability in the number and intensity of oligonucleotide spots on two-dimensional gel electrophoresis indicated that some degradation occurred during the RNA ligase reaction and subsequent isolation procedures. Quantitation of the ^{3}H-labeled oligonucleotides, however, was consistent over several experiments, and in agreement with the previous analysis of the cap structures of ovalbumin mRNA (1). While the ^{32}P-labeled

samples were the obvious choice for sequence analysis for the advantages mentioned, the relative abundance of the 5'-terminal oligonucleotides was determined using ^{3}H-labeled samples.

The proposed secondary structure near the 5' terminus of $mRNA_{ov}$ (2) is shown for the Ov 8 form.

$$m^{7}GpppACAUACAGCU^{A}G$$
$$3'....G^{U}{}^{UAUGUCGA}{}_{A}{}^{A}$$

The 5'-terminal nucleotide extends from the duplex part of the hairpin by two nucleotides for the Ov 8 form and four nucleotides for the Ov 10 form. The 5'-terminal nucleotide for Ov 6 would be part of the duplex. The variable position of the 5' terminus relative to the duplex could cause differences in the extent of post-transcriptional modifications such as capping or methylation. Since the $m^{7}G$ is important for initiating translation (25), any alteration in its accessibility to initiation factors could affect the rate of initiation. Thus it is possible that not all forms of $mRNA_{ov}$ are translated at the same rate.

The different termini of $mRNA_{ov}$ could arise from having more than one site for either initiating transcription or processing a primary transcript initiated elsewhere. Previous examples of 5'-terminal heterogeneity, found for late mRNAs of SV40 (26) and polyoma viruses (27) and the DNA binding protein mRNA of Adenovirus 2 (28), were attributed to the lack of a Goldberg-Hogness box, a putative promoter sequence of the general form, 5'-TATAAA-3'. Since the ovalbumin gene contains this sequence, the correlation of heterogeneity with the absence of a Goldberg-Hogness sequence would not seem to apply. More recently, 5'-terminal heterogeneity has been found for many of the transcription units of Adenovirus 2 that also contain this sequence (Dr. C. Baker, personal communication). Thus, the finding of heterogeneity at the 5' termini of eukaryotic mRNA may be a widespread phenomenon, unrelated to the presence of the Goldberg-Hogness sequence.

ACKNOWLEDGEMENTS

We are indebted to Drs. Pierre Chambon and Günther Schütz for generously providing bacterial strains containing plasmids for ovalbumin, conalbumin and ovomucoid. We are also grateful to Gary M. Hellmann and Beth Genya Cohen for helpful discussions, Robert S. Young for technical assistance, and Ms. Ginger Hodges for help with the manuscript. This work was supported by grants from the National Institute of General Medical Sciences (GM 20818) and the National Cancer Institute (CA 27801).

REFERENCES

1. Malek, L.T., Breter, H.-J., Hellmann, G.M., Friderici, K.H., Rottman, F.M. and Rhoads, R.E. (1979). J. Biol. Chem. 254, 10415-10420.
2. McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M., and Brownlee, G.G. (1978). Nature 273, 723-728.
3. Hughes, S.H., Payvar, F., Spector, D., Schimke, R.T., Robinson, H.L., Payne, G.S., Bishop, J.M., and Varmus, H.E. (1979). Cell 18, 347-359.
4. Chu, L.-Y., Lockard, R.E., RajBhandary, U.L., and Rhoads, R.E. (1978). J. Biol. Chem. 253, 5228-5231.
5. Clewell, D.B. (1972). J. Bacteriol. 110, 667-676.
6. Maniatis, T., Jeffrey, A., and Kleid, D.G. (1975). Proc. Natl. Acad. Sci. USA 72, 1184-1188.
7. Noyes, B.E., and Stark, G.R. (1975). Cell 5, 301-310.
8. Childs, G., Levy, S., and Kedes, L.H. (1979). Biochemistry 18, 208-213.
9. Adams, S.L., Alwine, J.C., Crombrugghe, B., and Pastan, I. (1979). J. Biol. Chem. 254, 4935-4938.
10. Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5350-5354,
11. Munns, T.W., Liszewski, M.K. and Sims, H.F. (1977). Biochemistry 16, 2163-2168.
12. Bruce, A.G., and Uhlenbeck, O.C. (1978). Nucleic Acids Res. 5, 3665-3677.
13. Breter, H.-J., Malek, L.T., Hellmann, G.M., and Rhoads, R.E. (1979). Analytical Biochem 98, 102-111.
14. DeWachter, R., and Fiers, W. (1972). Analytical Biochem. 49, 184-197.
15. Lockard, R.E., Alzner-Dweerd, B., Heckman, J.E., MacGee, J., Tabor, M. W., and RajBhandary, U.L. (1978). Nucleic Acids Res. 5, 37-56.
16. Efstradiatis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K., and Kafatos, F.C. (1977). Nucleic Acids Res. 12, 4165-4174.
17. Sanger, F., and Coulson, A.R. (1977). FEBS Lett. 87, 107-110.
18. Laskey, R.A., and Mills, A.D. (1975). Eur. J. Biochem. 56, 335-341.
19. Hynes, N.E., Groner, B., Sippel, A.E., Nguyen-Huu, M.C., and Shutz, G. (1977). Cell 11, 923-932.
20. Lee, D.C., McKnight, G.S., and Palmiter, R.D. (1978). J. Biol. Chem. 253, 3493-3503.
21. Malek, L.T. (1980) Doctoral Dissertation.
22. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P. (1979). Nature 278, 428-434.
23. O'Hare, K., Breathnach, R., Benoist, C., and Chambon, P. (1979). Nucleic Acid Res. 7, 321-334.
24. Wasylyk, B., Kedinger, C., Corden, J., Brison, O., and Chambon, P. (1980). Nature 285, 367-373.
25. Banerjee, A.K. (1980). Microbiol. Rev. 44, 175-205.
26. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P., and Weissman, S.M. (1978). J. Mol. Biol. 126, 813-846.
27. Flavell, A.J., Cowie, A., Legon, S., and Kamen, R. (1979). Cell 16, 357-371.
28. Baker, C.C., Herisse, J., Courtois, G., Galibert, F., and Ziff, E. (1979). Cell 18, 569-580.