Volume 5 Number 7 July 1978 Nucleic Acids Research

Nucleotide sequence at the 5' end of ovalbumin messenger RNA from chicken

Daniel Kuebbing and Charles D.Liarakos

Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA

Received 17 May 1978

ABSTRACT

DNA sequence analysis of 300 nucleotides from the region of cloned, double-stranded ovalbumin cDNA corresponding to the 5' end of ovalbumin messenger RNA was accomplished using the technique of Maxam and Gilbert (Proc. Nat. Acad. Sci. USA (1977) 74,560-564). The AUG initiation codon was located 52 nucleotides from the AT linkers used in cloning and immediately adjacent to the amino terminal peptide of ovalbumin, indicating the absence of a "signal peptide" in this protein. The nucleotide sequence coding for a phosphorylated peptide from ovalbumin was also found. These results demonstrate that the coding portion of $mRNA_{ov}$ begins near the 5' end of the molecule leaving some 600 nucleotides of non-coding information at the 3' end.

INTRODUCTION

The successful cloning of the double-stranded DNA copy of ovalbumin messenger RNA in plasmid pOV_{230} (1) has produced enough DNA to perform DNA sequence analysis by the chemical degradation procedure of Maxam and Gilbert (2). Using this technique, we have determined the nucleotide sequence of a contiguous series of restriction endonuclease fragments from the region of ovalbumin $cDNA_{ds}$ corresponding to the 5' end of ovalbumin mRNA (3). DNA sequence analysis of this region was considered necessary to unambiguously determine, (a) the location of both the AUG initiation codon and the 1161 nucleotide coding portion of $mRNA_{ov}$ within the 1890 nucleotide messenger RNA molecule; and (b) the presence or absence of a nucleotide sequence coding for a "signal peptide". The size and sequence of the 5' noncoding region of $mRNA_{ov}$ might also provide insight into understanding the translation of this mRNA, and, in studies of genomic ovalbumin DNA, would identify the location of possible, contiguous "control" sequences, thereby adding to our understanding of the mechanism by which

estrogen induces the synthesis of ovalbumin mRNA in the chick oviduct.

MATERIALS AND METHODS

γ-^{32}P-ATP was prepared by the method for Glynn and Chappell (4) as described by Maxam and Gilbert (2). Dimethylsulfate and piperidine were purchased from Aldrich and Fisher respectively. Hydrazine (95%) and imidazole were purchased from Eastman. All other chemicals were reagent grade.

Restriction endonucleases Hinf I, Sst I, Pst I and Taq I were purchased from Bethesda Research Labs. T4 polynucleotide kinase was purchased from PL Biochemicals or Boehringer Mannheim.

pOV_{230} DNA was purified as described by McReynolds et al. (1) and was supplied by Dr. A. Dugaiczyk. pMB9, the parent plasmid of pOV_{230}, was supplied by Dr. L. McReynolds.

Restriction endonuclease digestions. DNA was digested with restriction endonucleases in 6 mM Tris-HCl (pH 7.5) - 6 mM $MgCl_2$-6 mM 2-mercaptoethanol at 37°C for 3-5 hours. Digestion mixtures were supplemented with NaCl as follows; Pst I, 6 mM NaCl; Sst I, 90 mM NaCl; Hinf I, 20 mM NaCl. Taq I digestions were done at 50°C in 10 mM Tris-HCl (pH 8.4)-6 mM $MgCl_2$-6 mM 2-mercaptoethanol. For primary digestion of total plasmid, an enzyme to DNA ratio of 1-2 units/µg was used. For secondary digestions of ^{32}P-labeled fragments, this ratio was 5 units/µg DNA.

5' End labeling of restriction fragments. Double-stranded DNA restriction fragments were ^{32}P-labeled at the 5' end of each strand using the polynucleotide kinase exchange reaction described by Berkner and Folk (5). γ-^{32}P-ATP was evaporated to dryness in a siliconized glass tube with a stream of filtered air just prior to use. The final reaction mixture contained: 10 µM γ-^{32}P-ATP (1500-2000 Ci/mmole), 300 µM ADP, 45 mM KCl, 18 mM $MgCl_2$, 4.5 mM dithiothreitol, 25 mM imidizole (pH 6.6), 20-50 pmoles of DNA fragments, and 1 unit of T4 polynucleotide kinase per pmole of DNA fragments. The final reaction volume was adjusted to ten times the volume of polynucleotide kinase added. The reaction mixture was incubated 15 minutes at 37°C in a glass capillary and the resulting ^{32}P-labeled DNA fragments were immediately separated from residual γ-^{32}P-ATP by chromatography at 3°C on a 0.9 x 27 cm Sephadex G-50 column in 10 mM Tris-HCl (pH 7.5)-1 mM EDTA. The

void volume fractions were collected and the ^{32}P-labeled DNA precipitated at -20°C after addition of 1/20 volume of 5 M NaCl and 2 volumes of ethanol.

Polyacrylamide gel electrophoresis and recovery of DNA. DNA restriction fragments from primary or secondary digestions were separated by electrophoresis in 6% polyacrylamide slab gels (1:30, bis:acrylamide) containing 50 mM Tris-borate (pH 8.3) and 1mM EDTA. Unlabeled DNA bands in the gel were stained with ethidium bromide (1 μg/ml) and visualized under ultraviolet light. ^{32}P-labeled DNA fragments were detected by autoradiography using DuPont Cronex 4 X-ray film. Some ^{32}P-labeled DNA fragments were strand separated by heat denaturation (5 minutes at 90°C) in 50% formamide followed by rapid cooling to 0°C and electrophoresis (200-400 volts) in an 8% polyacrylamide gel containing 7M urea-50 mM Tris-borate (pH 8.3) -1 mM EDTA. Individual DNA bands were cut out of the gels and the DNA extracted using the "crush and soak" technique described by Maxam and Gilbert (2). Optimal DNA recovery (>90%) was obtained with a buffer to gel ratio (v/v) of at least 4:1.

DNA sequencing. DNA sequencing was performed essentially as described by Maxam and Gilbert (2). The dimethylsulfate cleavage reactions for adenine and guanine residues were carried out for 25 and 15 minutes respectively. In some cases the alternate A-procedure (alkaline cleavage) was used. Hydrazinolysis in the presence (C-specific) or absence (pyrimidine specific) of 5 M NaCl was carried out for 30 and 20 minutes respectively. The products of the individual cleavage reactions were separated by electrophoresis in 20% or 12% polyacrylamide slab gels containing 7 M urea, 50 mM Tris-borate (pH 8.3), and 1 mM EDTA. Two or three successive loads of the gel allowed the unambiguous reading of 60-90 nucleotides. If amino acid sequence data was available to resolve ambiguities, as many as 120 nucleotides could be determined. Each restriction fragment was sequenced at least twice to insure accuracy and establish overlaps.

RESULTS

Previous work in our laboratory (3) has demonstrated the relative location of several restriction endonuclease sites in double stranded ovalbumin cDNA. The enzyme Hinf I was of particular interest because it cut the ovalbumin $cDNA_{ds}$ twice in the region

expected to contain the AUG initiation codon. Furthermore, Monahan et al. (3) predicted that this enzyme might also cut in a region coding for a known ovalbumin peptide. We therefore began our sequence analysis at the HinfI sites in pOV_{230}, the cloned form of $cDNA_{ds}$. Comparison of the Hinf I digestion patterns of parent plasmid pMB9 and plasmid pOV_{230}, which contains the complete ovalbumin $cDNA_{ds}$, revealed five new, pOV_{230}-specific fragments (F2, F4, F6, F7, and F9) and the disappearance of one of the original pMB9 fragments (the second largest) (Fig. 1). This missing pMB9 fragment contains the EcoRI site at which ovalbumin $cDNA_{ds}$ was inserted into the plasmid. Furthermore, the ovalbumin cDNA is oriented in the plasmid such that the Hinf I fragment containing the 3' end of the cDNA will also contain 65 bp of pMB9 DNA while the Hinf I fragment containing the cDNA 5' end will contain at least 620 bp of pMB9 DNA (1,6). Since there is also an internal 800 bp Hinf I fragment in ovalbumin $cDNA_{ds}$ (3), pOV_{230} Hinf I fragments F4 and F2 were ordered by their size as shown in Figure 3. Digestion of ^{32}P-labeled fragments F6, F7 and F9 with restriction

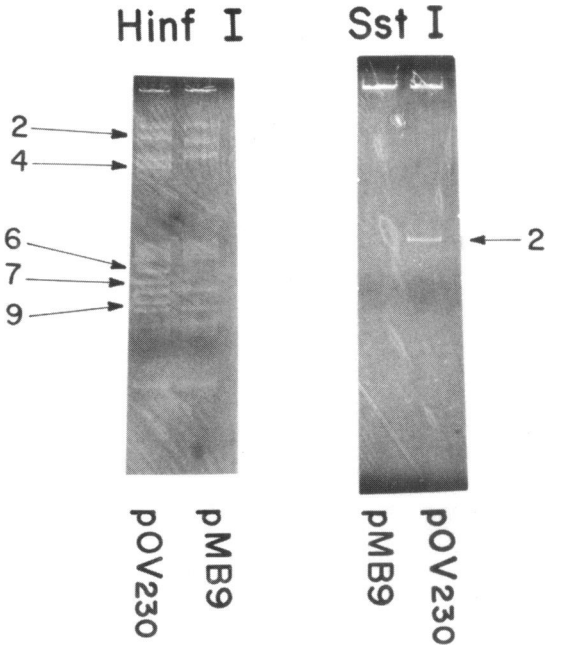

Fig. 1. Primary restriction endonuclease digestions. Hinf I digest of (a) pOV_{230} and (b) pMB9, and Sst I digest of (c) pMB9 and (d) pOV_{230} separated by 6% polyacrylamide gel electrophoresis and stained with ethidium bromide.

endonuclease Pst I revealed that only fragment F7 was cut (Fig.2). Since the only Pst I site in ovalbumin $cDNA_{ds}$ is approximately 450 bp from the 3' end of the coding strand (3), location of fragment F7 on ovalbumin $cDNA_{ds}$ was determined (Fig. 3). Previous results (3) also indicated that restriction endonuclease Sst I cut ovalbumin $cDNA_{ds}$ only once approximately 400 bp from the 3' end of the cDNA strand; however, Sst I digestion of pOV_{230} gave two fragments (Fig. 1)-one 400 bp long and the other too large to enter into the 6 % polyacrylamide gel. Since pMB9 contains no Sst I sites, this result indicated the presence of a second Sst I site, very close to the 3' end of the cDNA strand, in ovalbumin $cDNA_{ds}$. In

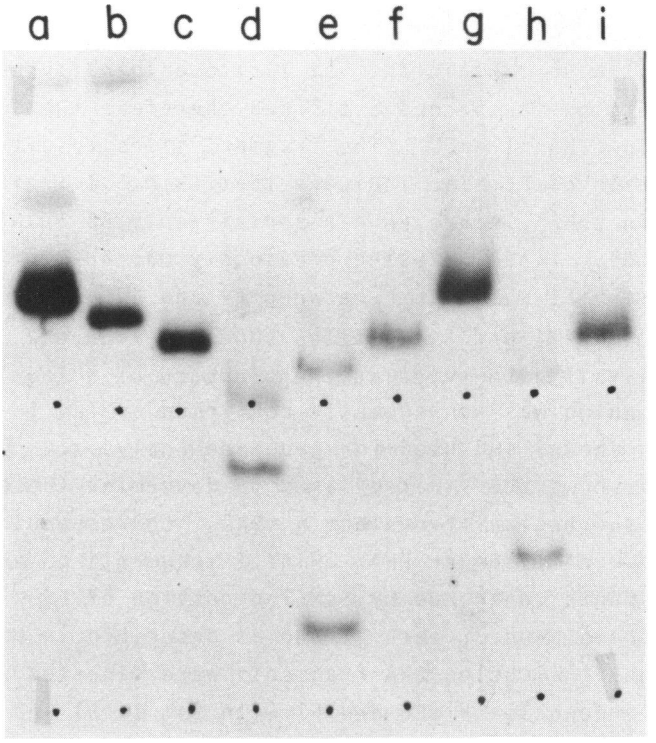

Fig. 2. Secondary digestion of Hinf I restriction fragments from pOV_{230}. Fragments F6, F7, and F9 were 5' end labeled with ^{32}P, digested with Sst or Pst and electrophoresed on a 6% polyacrylamide slab gel. (a) F6, (b) F7, (c) F9, (d) F6 + Sst I, (e) F7 + Sst I, (f) F9 + Sst I, (g) F6 + Pst I, (h) F7 + Pst I, (i) F9 + Pst I. The black dots indicate the position of the marker dyes, xylene cyanol (=200N) and bromphenol blue (=35N).

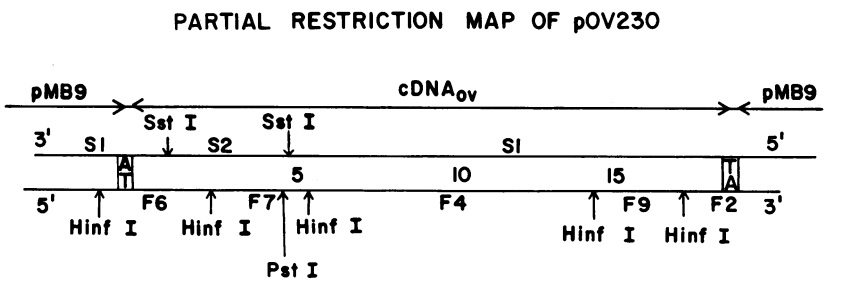

Fig. 3. Restriction map of pOV_{230} illustrating the relative positions of the Hinf I and Sst I fragments. Length in nucleotides x 10^{-2} is shown between the DNA strands.

addition, both Hinf I fragments F6 and F7, but not F9, were found to be cut by Sst I (Fig.2). Since the location of F7 was established by Pst I digestion, it must also contain the Sst I site originally reported (3), while F6 contains the second Sst I and therefore the 3' end of the ovalbumin cDNA strand. (The fragment sizes resulting from Sst I digestion of F6 also indicate that this Sst I site is in the ovalbumin $cDNA_{ds}$ since F6 can contain only 65 bp of pMB9 DNA at its 3' end). Finally, using previously mapped Hinf I sites (3,7), fragment F9 was located between F4 and F2 giving the relative placement of Hinf I and Sst I sites shown in Figure 3. Initial sequence analysis also revealed the presence of a Taq I site in fragment F6 which was subsequently confirmed by Taq I digestion of F6 (data not shown) and used for sequence analysis.

Figure 4 shows the strategy used to determine the entire sequence of the region of ovalbumin $cDNA_{ds}$ corresponding to the 5' end of ovalbumin messenger RNA. Hinf I fragments F6 and F7 as well as fragments generated by Sst I digestion of F6 were ^{32}P-labeled at the 5' end of each strand as described in Materials and Methods. The ^{32}P-labeled DNA fragments were digested with a second restriction endonuclease (F6 and F7 with Sst I; S1 and S2 with Hinf I; and the Taq I fragments from F6 with Sst I) and the resulting subfragments, each labeled at only one 5' end, were separated by polyacrylamide gel electrophoresis as described above. In the case of the F6 subfragment extending from the Hinf I site in pMB9 through the A-T linkers to the Taq I site in ovalbumin $cDNA_{ds}$, the two ^{32}P-labeled 5' ends were isolated by strand separation as described above. Figure 5 shows the gel from which the nucleotide

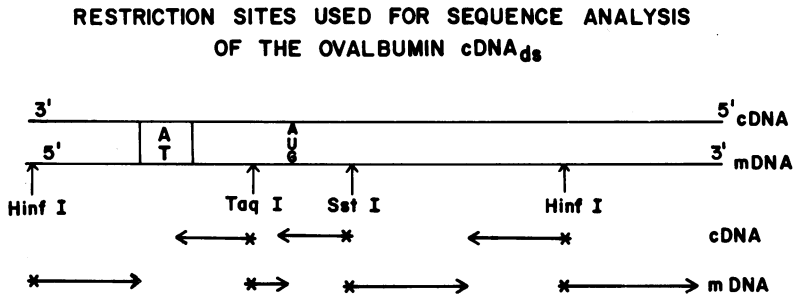

Fig. 4. Expanded restriction map of pOV_{230} showing the restriction sites used for sequence analysis of the region of pOV_{230} corresponding to the 5'end of $mRNA_{ov}$. The arrows indicate the strands (cDNA or mDNA) that were sequenced from their ^{32}P-labeled, 5' ends (*).

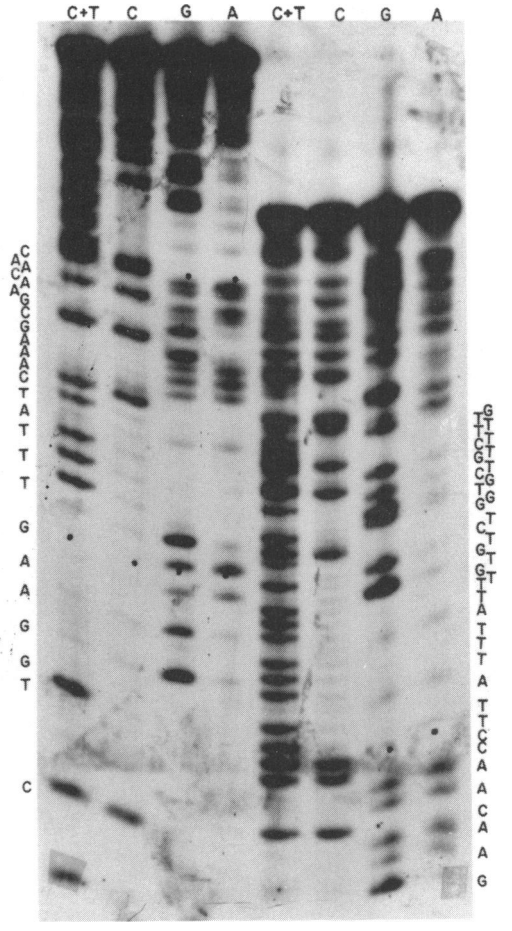

Fig. 5. DNA sequence gel from the first ovalbumin $cDNA_{ds}$ Hinf I site in pOV_{230}. This sequence reads in the cDNA strand from the Hinf I site back towards the first Sst I site.

sequence from the first Hinf I site in the ovalbumin $cDNA_{ds}$ back towards the first Sst I site was determined. As suggested previously (3), a known phosphorylated peptide from ovalbumin (8) was encoded in the location of this first Hinf I site. Figures 6 and 7 show the gels from which the sequence around the AUG initiation codon (as determined from the Sst I and Taq I sites respectively) was determined. The first AUG codon lies 52 nucleotides from the A-T linkers used in cloning and directly adjacent to the amino terminal peptide of ovalbumin. There is no intervening "signal

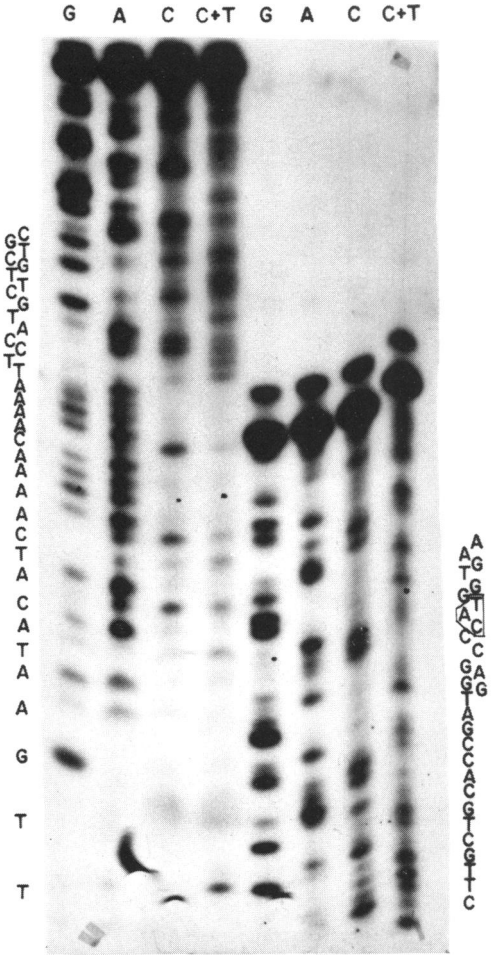

Fig. 6. DNA sequence gel from the first Sst I site in pOV_{230}. This sequence reads in the cDNA strand from the first Sst I site back towards the AUG initiation codon. The sequence complementary to the initiation codon (CAT) is also shown (box).

sequence" between this AUG codon and the amino terminal glycine of ovalbumin. Finally, Figure 8 illustrates the sequence of 302 nucleotides of ovalbumin $cDNA_{ds}$ corresponding to the 5' end of ovalbumin messenger RNA. The nucleotides are numbered using the convention suggested by Efstratiadis et al. (9); that is, the first nucleotide after the initiation codon is number one. The underlined amino acids represent known peptides that were used in confirming the nucleotide sequence.

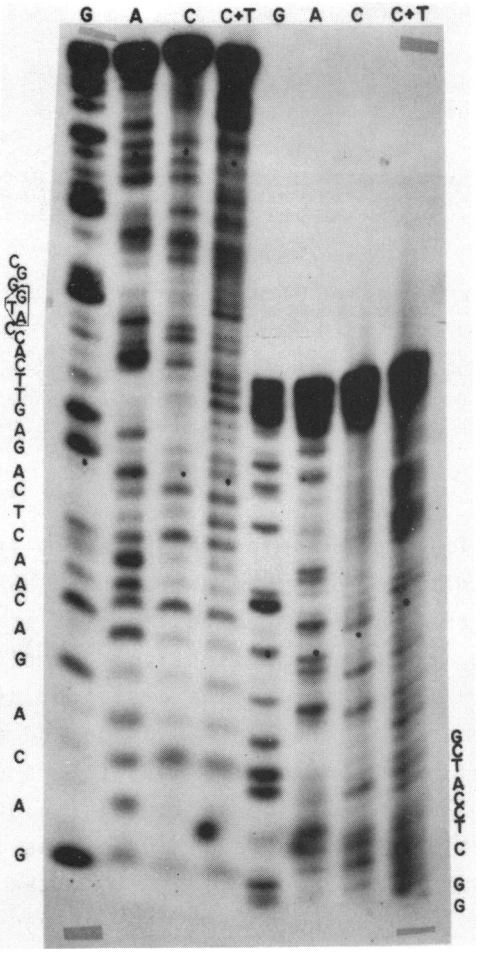

Fig. 7. DNA sequence gel from the Taq I site in pOV230. This sequence reads in the mDNA strand from the Taq I site towards the AUG initiation codon. The initiation codon appears as ATG (box).

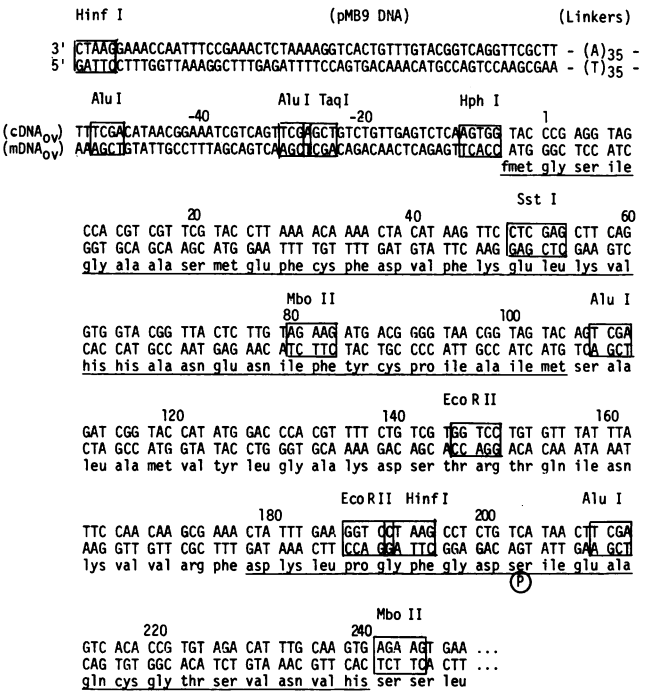

Fig. 8. Nucleotide sequence of ovalbumin $cDNA_{ds}$ corresponding to the 5' end of $mRNA_{OV}$. Also included are 58 nucleotides from the plasmid pMB9. Known peptides from ovalbumin are underlined.

DISCUSSION

From the size of the ovalbumin $cDNA_{ds}$ insert in pOV_{230} and its ability, when hybridized to ovalbumin messenger RNA, to completely protect the mRNA from S1 nuclease digestion, it is possible to conclude that pOV_{230} contains an essentially complete copy of the ovalbumin mRNA (1). The 52 nucleotide distance, as determined by nucleotide sequence analysis, of an AUG codon and adjacent, ovalbumin amino terminal peptide from the A-T linkers used in cloning indicates that the 1161 nucleotide coding region of ovalbumin mRNA begins relatively close to the 5' end of the molecule. Since ovalbumin mRNA contains 1890 nucleotides (10), the 3' end of this mRNA must contain more than 600 nucleotides of non-coding information. The function of this large, 3' non-coding region is presently unkown, but it has been shown to have some limited sequence homology to the 3' non-coding region of rabbit globin mRNA (11).

The first AUG codon in ovalbumin $cDNA_{ds}$ immediately precedes

the codon for the amino terminal glycine of ovalbumin. This result strongly suggests that this AUG is the true initiation codon and, therefore, that ovalbumin, unlike many other secretory proteins, does not contain a "signal peptide". Palmiter et al. (12) have drawn similar conclusions by determining the amino acid sequence of the 35 amino terminal residues of ovalbumin (See Fig. 8) synthesized by the in vitro translation of ovalbumin mRNA. These investigators found that the primary translation product began with methionine followed immediately by the amino terminal peptide of ovalbumin and that the sequence of this peptide was identical to that of secreted ovalbumin. Most secretory proteins, including the other egg white proteins conalbumin, ovomucoid, and lysozyme (13,14) have been shown to be translated as pre-proteins which contain approximately 20 extra amino acids, most (~80%) of which are hydrophobic, at their amino terminal ends. This extra peptide is believed to be necessary for the attachment of the ribosomes synthesizing these proteins to the endoplasmic reticulum thereby facilitating the vectorial transport of these proteins through the membrane reaulting in their accumulation in the Golgi apparatus and eventual secretion (15). The signal peptide is usually removed during membrane transport. Inspection of the amino terminal region of ovalbumin reveals that 12 of the first 20 amino acids (60%) are hydrophobic. It is possible, therefore, that this region of the protein serves as a "signal sequence" which remains an integral part of the mature protein. Alternatively, the acetylation of the amino terminal glycine residue early in the synthesis of ovalbumin may play a role in its secretion (12).

Examination of the nucleotide sequence around the AUG initiation codon reveals a possible hairpin loop structure (Figure 9, structure A) containing eight base pairs and including the entire AUG codon. Using Tinoco's revised rules (16), the free energy of formation of this hairpin is estimated at -6.4 kcal, suggesting a stable structure. Figure 9 also illustrates two possible regions of complementarity between the 5' end of ovalbumin mRNA and the highly conserved sequence at the 3' end of 18S ribosomal RNA (17,18). In "structure A", nucleotides from the mRNA hairpin loop are not involved in binding to 18S rRNA; whereas, in "structure B", the mRNA hairpin is reduced to a less stable, four base pair

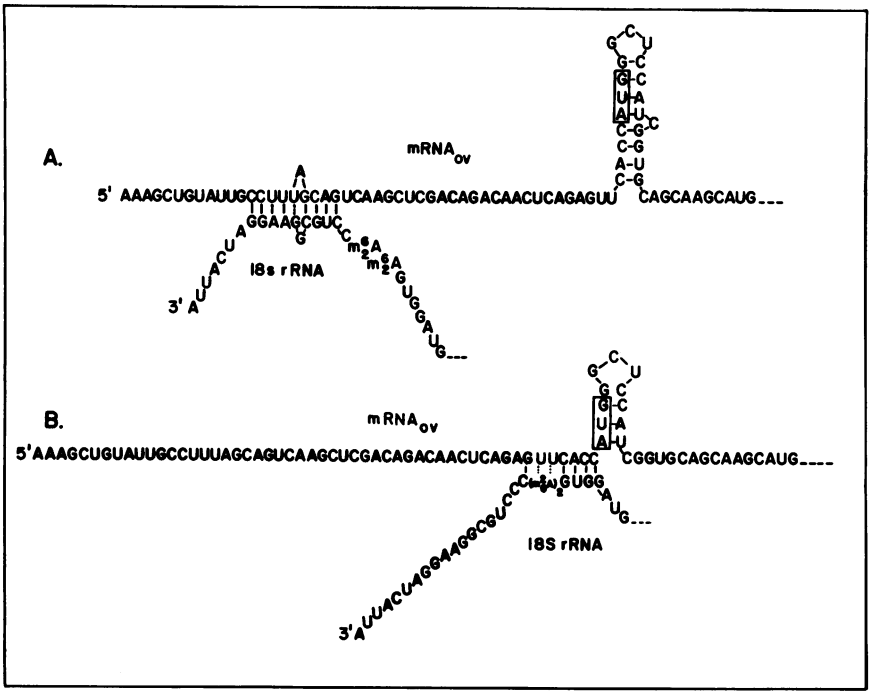

Fig. 9. Nucleotide sequence at 5' end of ovalbumin mRNA showing a possible hairpin loop containing the AUG initiation codon (structure A) and possible, complementary base pairing with the 3' end of 18S rRNA (structures A and B). In structure B., the two m^6_2A-U base pairs may be similar to G-U base pairs in their hydrogen bonding or may not form at all.

structure (net free energy = -1.0 kcal) by binding to the 18S rRNA. Similar possible base pairing has been suggested between the 3' end of 18S rRNA and the non-coding, 5' ends of several eucaryotic and viral messenger RNAs (18), however, such interactions are speculative and their role, if any, in the initiation of protein synthesis remains to be determined.

Recent evidence (19,20,21) has indicated that the coding region of the ovalbumin gene is interrupted at several points by regions of non-coding DNA of varying lengths. Preliminary DNA sequence analysis, in our laboratory, at the Sst I and Hinf I sites in genomic ovalbumin DNA has verified the presence of several regions of non-coding, "insert" DNA, indicating a far more complicated structure than is represented by ovalbumin $cDNA_{ds}$. (This data will be reported elsewhere.)

Finally, using an mRNA template-cDNA primer extension method,

McReynolds et al. (22) have recently demonstrated the presence of twelve additional nucleotides at the 5' end of ovalbumin mRNA not contained in pOV_{230} DNA. This region was sequenced and found to contain no AUG codon, but was capable of forming an eight base pair, hairpin loop at the 5' end of the mRNA. The sequence data of these investigators are in essential agreement with the sequence reported here.

ACKNOWLEDGEMENTS

We gratefully acknowledge Drs. Bert O'Malley and Jeffrey Rosen for their critical review of this manuscript and Dr. L. McReynolds and G. Brownlee for providing a copy of their manuscript prior to publication. This work was supported under NIH grants HD08188-06 and HD07495-05. D.K. is a post-doctoral fellow supported under NIH training grant HD0444-04.

ABREVIATIONS

$mRNA_{ov}$ = ovalbumin messenger RNA
cDNA = the complementary DNA copy of a mRNA
$cDNA_{ds}$ - double-stranded cDNA
mDNA = the homologous DNA copy of a mRNA

REFERENCES

1 McReynolds, L.A., Catterall, J.F. and O'Malley, B.W (1977) Gene 2, 217-231.
2 Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
3 Monahan, J.J., Woo, S.L.C., Liarakos, C.D. and O'Malley, B.W. (1977) J. Biol. Chem. 252, 4722-4728.
4 Glynn, I.M. and Chappell, J.B. (1964) Biochem J. 90, 147-149.
5 Berkner, K.L. and Folk, W.R. (1977) J. Biol. Chem. 252, 3176-3184.
6 Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976) Cell 8, 163-182.
7 Brownlee G.G. and Cartwright, E.M. (1977) J. Mol. Biol. 114, 93-117.
8 Milstein, C.P. (1968) Biochem. J. 110, 127-134.
9 Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977) Cell 10, 571-585.
10 Woo, S.L.C., Rosen, J.M., Liarakos, C.D., Choi, Y.C., Bu ch, H., Means, A.R. and O'Malley, B.W. (1975) J. Biol. Chem. 250, 7027-7039.
11 Cheng, C.D., Brownlee, G.G., Carey, N.H., Doel, M.T., Gillam, S. and Smith, M. (1976) J. Mol. Biol. 107, 527-547.
12 Palimiter, R.D., Gagnon, J. and Walsh, K.A. (1978) Proc. Nat. Acad. Sci. USA 75, 94-98.
13 Palmiter, R.D., Gagnon, J., Ericsson, L.H. and Walsh, K.A. (1977) J. Biol. Chem. 252, 6386-6393.

14 Strauss, A.W., Bennett, C.D., Donohue, A.M., Rodkey, J.A. and Alberts, A.W. (1977) J. Biol. Chem. 252, 6846-6855.
15 Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67, 835-851.
16 Tinoco, I., Borer, P.N., Dengler, B., Levine, M., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
17 Alberty, H. Raba, M. and Gross, H.J. (1978) Nucleic Acids Res. 5, 425-434.
18 Hagenbuchle, O., Santer, M. and Steitz, J.A. (1978) Cell 13, 551-563.
19 Breathnach, R., Mandel, J.L. and Chambon, P. (1977) Nature 270, 314-319.
20 Doel, M.T., Houghton, M., Cook, E.A. and Carey, N.H. (1977) Nucleic Acids Res. 11, 3701-3713.
21 Lai, E.C., Woo, S.L.C., Dugaiczyk, A., Catterall, J.F. and O'Malley, B.W., (1978) Proc. Nat. Acad. Sci. USA, in press.
22 McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G., Personal communication.