Heterogeneous initiation regions for transcription of the chicken ovomucoid gene

Eugene C.Lai, Dennis R.Roop*, Ming-Jer Tsai, Savio L.C.Woo and Bert W.O'Malley

Department of Cell Biology and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

Received 22 February 1982; Revised and Accepted 20 August 1982

ABSTRACT

We have identified the 5' termini of ovomucoid RNA transcripts isolated from estrogen stimulated chick oviduct by S1-nuclease mapping. Most of the ovomucoid RNA molecules (88-89%) are initiated at a predominant region which is positioned at about 30 basepairs downstream from a "TATATAT" sequence. However, there is at least one additional initiation region which accounts for 9% of the ovomucoid RNA transcripts. It is located at approximately 80 basepairs further upstream (5') from the major initiation region. DNA sequence data reveal that an AT-rich sequence resembling the "TATA" box is also present 30 nucleotides before this minor region. Our finding provides additional insight into the molecular mechanism involved in the transcription initiation of eucaryotic genes, and indicates that *in vivo* synthesis of mRNA may not require an absolute site for initiation of transcription.

INTRODUCTION

Ovomucoid is a glycoprotein which accounts for about 10% of the egg-white protein in chicken (1). Its synthesis in the oviduct is regulated by the steroid hormones, estrogen and progesterone (2,3). During acute estrogen stimulation of withdrawn chicks, a marked increase in the number of ovomucoid mRNA molecules occurs first in the nucleus and subsequently in the cytoplasm of oviduct cells; and this event precedes the accumulation of the ovomucoid protein in the cytoplasm (4,5). Recent studies have indicated that the steroid-mediated induction of ovomucoid synthesis in the chick oviduct is via a "transcriptional control" mechanism (5). In order to further understand the molecular mechanism involved in the expression and regulation of this protein in the chicken oviduct, we have cloned the entire ovomucoid natural gene from the chicken genome (6,7). Results from restriction endonuclease mapping, electron microscopy and direct DNA sequencing have provided detailed information concerning the molecular structure of the ovomucoid gene (7). Using specific DNA fragments of the cloned gene as hybridization probes, we have mapped the 5' transcription domain of the ovomucoid gene. We report that ovomucoid mRNA molecules isolated from stimulated chick oviduct are heterogeneous at their 5' ends.

MATERIALS AND METHODS

Preparation of Hybridization Probes for S1 Nuclease Mapping

pOM4.0 is a chimeric plasmid containing a 4.0 Kb Hind III-cleaved chicken DNA insert which consists of approximately 0.25 Kb of the 5' region of the ovomucoid gene and 3.75 Kb of the 5' flanking sequence (Fig. 1). The restriction map of this DNA has been determined previously (7). Three hybridization probes were prepared from pOM4.0 as follows. Fragment A was obtained by digesting 200 µg of the recombinant plasmid with Hpa II and Bgl II (BRL) until completion using recommended conditions, and by subjecting the digest to electrophoresis in a 4% polyacrylamide gel using the Tris-borate buffer. After staining with ethidium bromide, the 1.7 Kb fragment was located and excised from the gel, homogenized using a Tekmar tissumizer and resuspended in 2 volumes of 10 mM Tris-HCl, pH 7.6 and 1 mM EDTA (TE). DNA was allowed to defuse out of the gel by gently rocking the mixture overnight at room temperature. The solution was passed through two layers of glass fiber filter and precipitated with ethanol. Fragment B (1.67 Kb) was obtained by digesting pOM4.0 with Ava I and Bgl II and extracted as described above. Fragment C (1.82 Kb) was similarly obtained except pOM4.0 was digested with Hind III and Bgl II.

S1 Nuclease Mapping

The isolated DNA fragments (Fragment A, B and C, respectively) were labeled at their 5' ends by T4 polynucleotide kinase using γ -[³²P]ATP (Amersham) as described previously (7). The labeled DNA was digested with Taq I, and the fragment which was labeling at the 5' end of the coding strand for ovomucoid was extracted after a 4% polyacrylamide gel electrophoresis. DNA-RNA hybridizations were performed in a final volume of 100 µl and contained 60% formamide (Fluka) 0.3 M NaCl, 0.04 M Pipes, pH 6.4, 0.001 M EDTA, approximately 20 ng of end-labeled probe (50,000 cpm) and 50 µg of oviduct poly(A) RNA. The reaction mixtures were denatured at 100°C for 5 min and incubated at 37° for 20 hrs. The hybrids were digested with S1 nuclease (2000 units, Miles) in a final volume of 500 µl which contained 0.2 M Na acetate, pH 4.5, 0.4 M NaCl, 0.002 M ZnCl₂ and 10 µg denatured calf thymus DNA. The reaction was incubated at 37° for 30 min. The digestion was terminated by the addition of 2 volumes of ethanol. After 10 min at -70°, the samples were centrifuged and the pellets were resuspended in 0.2 N NaOH. After 10 min incubation at 68°, the samples were diluted with an equal volume of 10 M urea, 0.05% xylene cyanol and 0.05% bromophenol blue and incubated for an additional 5 min at 68°. The samples were then layered on a 10% polyacrylamide gel containing 7M urea and subjected to electrophoresis at 150 V for 16 hr. Autoradiography with an intensifying

screen was performed at -20° for 1-2 days. Hormone treatment of White Leghorn chicks, isolation of nuclei, and isolation of nuclear RNA were described previously (8). The isolation of oviduct poly(A) RNA was by oligothymidylic acid-cellulose chromatography (9).

Densitometer Scanning

Autoradiograms of [^{32}P]labeled S1 nuclease products were scanned using the Quick Scanner (Helena Laboratories). The relative amounts of the [^{32}P]DNA fragments were quantitated by cutting out the area under each individual peak and weighing it with a Mettler electronic balance.

Nucleic Acid Electrophoresis, Transfer and Hybridization

RNA or DNA samples were suspended in 4 mM methylmercury hydroxide and then subjected to electrophoresis on an 1.5% agarose slab gel containing 4 mM methylmercury hydroxide as described by Bailey and Davidson (10). Immediately after electrophoresis, methylmercury hydroxide was removed from the gel, and the nucleic acids were partially degraded by alkali and transferred by blotting to freshly prepared diazobenzoyloxymethyl (DBM) paper as described by Alwine et al. (11). The nucleic acids bound to the DBM paper were hybridized to DNA probes, labeled with [^{32}P] by nick translation (specific activity 5×10^8 cpm per μg), washed and autoradiographed.

DNA Sequencing

The Hpa II-Bgl II fragment from pOM4.0 (see Figure 1) was labeled at the 5' ends with T4 polynucleotide kinase in the presence of γ -[^{32}P]ATP as described above. The labeled DNA was recut with Dde I so that a 225bp fragment labeled only at the coding strand could be isolated after polyacrylamide gel electrophoresis. This fragment was degraded chemically according to the method of Maxam and Gilbert (12) and analyzed by electrophoresis on a 20% urea-acrylamide gel (0.3 mm) according to Sanger and Coulson (13).

Mung Bean Nuclease Mapping

The [^{32}P]-labeled DdeI-HpaII DNA fragment was denatured and allowed to hybridize to chicken oviduct mRNA as described in S1 nuclease mapping. After incubation the hybrids were digested with 900 units of Mung bean nuclease (PL Biochemicals) at 37° for 1 hour in 500 μl of 0.1 M Na acetate, pH 4.5, 0.2M NaCl and 0.0001M ZnCl_2 . The remaining [^{32}P]-DNA was treated and analyzed for electrophoresis as described for S1 nuclease mapping.

RESULTS

S1 Nuclease Mapping of the 5' Ends of Ovomuroid Gene Transcripts

A chicken DNA fragment containing the 5' flanking sequence and the 5'

terminal region of the ovomucoid gene has been subcloned in pBR322. A structural map of this recombinant (pOM4.0) is shown in Fig. 1. The first structural sequence (SS 1) of the ovomucoid gene codes for the leader sequence and part of the protein coding sequence of the mature ovomucoid mRNA (7). The junction between SS 1 and the first intervening sequence of the gene (IVS A) has been located precisely by comparing the DNA sequences obtained from the cloned cDNA and the natural gene (14,15). Also, by comparing DNA sequences of the two clones we have tentatively identified the transcription initiation site for the ovomucoid mRNA (mRNA_{OM}) to be located at nucleotide 1 in Figure 1. In order to vigorously demonstrate that this is the authentic 5' terminus of the ovomucoid gene, it was required that the 5' end of ovomucoid mRNA synthesized in vivo be determined. Total poly(A)-containing RNA was isolated from estrogen stimulated chick oviduct and analyzed by S1 nuclease mapping as

Figure 1. Structure of the 5' terminal region of the chicken ovomucoid gene. A DNA sequence of 260 nucleotides starting from the 5' flanking sequence and extending to the Hind IIIb site of the ovomucoid gene is also presented. Ovomucoid structural sequence, ; ovomucoid intervening sequence, ; ovomucoid 5' flanking sequence, ; pBR322, ; transcription initiation site, ▲; protected fragment of the DNA probe (see Figure 2), TTTT*. Vertical arrows denote restriction sites.

developed by Berk and Sharp (16).

A specific hybridization probe was prepared from pOM4.0 as follows. This recombinant plasmid was digested with Bgl II and Hpa II, and a DNA fragment of 1.7 Kb in size was isolated by polyacrylamide gel electrophoresis. This fragment (Fragment A) extends from the Hpa II site in the first structural sequence of the ovomucoid gene into its flanking sequence and terminates at the Bgl II site 1.7 Kb upstream (Fig. 1). The isolated fragment was labeled at the 5' termini with γ -[32 P]ATP and further cleaved with Taq I. The Taq I-Hpa II fragment which was labeled only at the Hpa II terminus of the coding strand was again subjected to polyacrylamide gel electrophoresis and purified. This DNA was thermally denatured and allowed to hybridize with total oviduct poly(A)-containing RNA in buffer containing 60% formamide. The hybrid was treated with S1 nuclease and analyzed by polyacrylamide gel electrophoresis in 7M urea followed by autoradiography. Since the Hpa II site is located in the structural sequence of the ovomucoid gene, mRNA_{OM} will protect the labeled 5' terminus of this probe from nuclease digestion, and the size of the protected fragment reflects the distance between the 5' terminus of the mRNA_{OM} and the Hpa II cleavage site at nucleotide 20 (Fig. 1). Indeed, a group of [32 P]-labeled structural gene fragments, about 20 nucleotides in size, was obtained (Fig. 2A, lane a). However, an additional fragment of approximately 104 nucleotides in length was apparently also protected from S1 nuclease digestion, although it was present in a much reduced quantity as compared to the smaller fragments (Fig. 2A, lane a). The protected fragments were constant in size and always appeared in the autoradiogram in similar ratios even when the DNA-RNA hybrids were digested with increasing amounts of S1 nuclease (data not shown), indicating that the fragments were not generated as artifacts in the S1 nuclease treatment. Furthermore, no protected fragment was observed when the DNA probe was allowed to hybridize with yeast RNA (Figure 2A, lane c).

RNA molecules which are capped at the 5' end can often protect an additional 1-5 nucleotides of the DNA probe from digestion by S1 nuclease (17,18). Furthermore, the group of fragments of 18-24 nucleotides in length could be the result of minor variations at the 5' end of capped ovomucoid RNA, as such has been found to exist in chicken ovalbumin mRNA (19,20). Keeping the above considerations in mind, our S1 mapping results appear to indicate that the 5' termini of the majority of *in vivo* ovomucoid gene transcripts are located at about 20 nucleotides upstream from the Hpa II sites, i.e., around nucleotide 1 in Figure 1 as we have previously predicted. However, the presence of a larger S1-resistant fragment in our mapping experiment suggests that the 5' terminus of the ovomu-

Figure 2. Mapping of the 5' ends of ovomucoid gene transcripts. Panel A: Autoradiogram of [^{32}P]-labeled protected fragments arising from the hybridization of a) oviduct poly(A) RNA (30 μg) and Hpa II-Taq I probe; b) oviduct poly(A) RNA (30 μg) and Ava I-Taq I probe; c) yeast RNA (30 μg) and Hpa II Taq I probe d) yeast RNA (30 μg) and Ava I-Taq I probe. Lane M, [^{32}P]-labeled Hinf I digested ϕX174 DNA as size marker. Panel B: Autoradiogram of [^{32}P]-labeled protected fragments arising from the hybridization of a) yeast RNA and Hind III-Taq I probe; b) oviduct poly(A) nuclear RNA (100 μg) and Hind III-Taq I probe; Lane M, [^{32}P]-labeled Hae III digested pBR322 DNA as size marker. Panel C: tracing of the relative intensity of the radioactive bands shown in Panel B, lane b.

coid mRNA is not unique. In vivo transcription may also be initiated from another region approximately 84 nucleotides further upstream from the major initiation site.

In order to substantiate the authenticity of the larger protected fragment, a second hybridization probe was prepared from pOM4.0. This recombinant plasmid was digested with Bgl II and Ava I, and a 1.67 Kb DNA fragment (Fragment B) was extracted after gel electrophoresis. This fragment was labeled with [^{32}P] at the 5' termini and recut with Taq I. A 1.27 Kb Taq I-Ava I fragment labeled only at the Ava I site was purified and used as the hybridization probe. The Ava I site at the 5' region of the ovomucoid gene is located at 27 nucleotides upstream from Hpa II site, and therefore, 7 nucleotides beyond the 5' terminus of the predominant species of ovomucoid mRNA (Fig. 1). The labeled 5' terminus of this probe will be protected only by mRNA initiated upstream from the Ava I

site. If there are ovomucoid mRNA molecules whose 5' terminus begins at about 104 nucleotides from the Hpa II site, as revealed in the previous experiment, S1 nuclease mapping using the Taq I-Ava I probe should reveal a S1 resistant fragment shorter by 27 nucleotides (the distance between Hpa II and Ava I sites). Indeed, a single fragment of 77 nucleotides was found to be protected from S1 nuclease digestion, while the dominant group of fragments of 20-24 nucleotides in length was no longer present (Fig. 2A lane b).

A third probe for S1 nuclease mapping was generated by cleaving pOM4.0 with Bgl II and Hind III, and labeled with [32 P] at the 5' ends (Fragment C, Fig. 1). After further digestion with Taq I, the Taq I-Hind III fragment (which is only labeled at the 5' end of the coding strand and is about 1.42 Kb in length, Figure 1) was isolated for hybridization with oviduct nuclear poly(A) RNA in an S1 nuclease mapping experiment. Since the Hind III site is located within the first intervening sequence of the ovomucoid gene, only ovomucoid precursor RNA molecules will protect the [32 P] label while mature mRNA_{OM} will not. Two protected fragments were detected, the predominant one at about 141 nucleotides and a larger fragment at 225 nucleotides (Fig. 2B, lane b). The size differences of the fragments are again exactly those predicted from the distance between the Hpa II site and the Hind III site in the ovomucoid gene. This result again demonstrates heterogeneity in the 5' termini of the ovomucoid RNA whose locations are mapped to the same sites in the gene as detected by the previous two probes.

In order to quantitate the relative amounts of these two ovomucoid RNA species, the relative intensity of the bands was determined by densitometer scanning (Figure 2C). When the area under each of the traced peaks was determined and compared, it was calculated that the 141 nucleotide band is about 10 times more intense than the 225-nucleotide band. Assuming the hybridization rates of the ovomucoid RNA to the two protected fragments are equal since their sizes are quite close and exceed a hundred nucleotides, we estimate that about 89% of the ovomucoid RNA initiates at the major region, while 9% of it has a longer 5' terminus, and its transcription is initiated at about 84 nucleotides further upstream. The ovomucoid RNA we analysed were only nuclear precursors which contained at least the first intron. The relative ratio of these molecules is representative of the entire population of ovomucoid RNA only if we assume that the rates of processing for all the different species are the same. Because the rates of molecular hybridization will be considerably different if the protected fragments differ greatly (such as the 104- and 20- nucleotide fragments generated by the Hpa II-Taq I probe), there is presently no other suitable DNA probe or

method available to accurately quantitate the different species of ovomucoid transcripts in the oviduct cytoplasmic RNA.

Identification of Heterogeneous Species of Ovomucoid Transcripts After Gel Electrophoresis

Heterogeneity of transcripts of the ovomucoid gene also were confirmed by hybridizing [^{32}P]-labeled DNA probes to oviduct RNA which has been subjected to denaturing agarose gel electrophoresis and blotted onto diazobenzylxymethyl (DBM) paper. Oviduct nuclear RNA (100 μg) and cytoplasmic RNA (50 μg) were subjected to agarose gel electrophoresis in methylmercury hydroxide and stained in ethidium bromide (Figure 3, panel A). After the RNAs were transferred to DBM-paper, they were allowed to hybridize with the Ava I-Bgl II DNA probe (Fragment B) from pom4.0 (Fig. 1) which had been labeled with [^{32}P] by nick-translation. A hybridization signal of size 1100 nucleotides was clearly visible with both oviduct cytoplasmic and nuclear RNA (Fig. 3B, lane a and b). Furthermore, an additional 1700 nucleotide signal showed up in oviduct nuclear RNA (Fig. 3B, lane b.) This result indicates that oviduct RNA does contain transcripts which are coded by gene sequence upstream from the Ava I site in the region of the ovomucoid gene. That these hybridization signals were not generated by some contaminating ovomucoid structural sequence is demonstrated by the inability of the probe to hybridize with a 0.9 Kb DNA fragment containing the 5' half of a cloned ovomucoid cDNA (Fig. 3B, lane c). This DNA fragment was obtained by digesting pom502 (7) with Hha I and contained 700 nucleotides of ovomucoid structural sequence starting from nucleotide 1 in Figure 1. Using the ovomucoid cDNA as hybridization probe, multiple discrete bands containing ovomucoid structural sequences were observed in oviduct nuclear RNA (Fig. 3C) and these had been determined previously to be ovomucoid mRNA precursors at various stages of processing (21). The two most intense signals, which correspond to mature mRNA_{OM} (1100 nucleotides) and its most abundant species of higher molecular weight precursor RNA (1700 nucleotides), respectively, are identical to the hybridization signals generated by the Taq I-Ava I DNA probe. In our electrophoresis system, the heterogeneous species of ovomucoid RNA cannot be resolved into distinct bands because their sizes differ by only 84 nucleotides. Furthermore, any difference in size would be obscured by the heterogeneous length of poly(A) nucleotides at the 3' end of the RNA.

Localization of the Major and Minor Initiation Regions for the Ovomucoid Transcripts

The precise positions of the 5' ends of the two species of ovomucoid RNA can be determined using sequencing gel-electrophoresis. A 225 bp DdeI-Hpa II

Figure 3. Identification of heterogeneous species of ovomucoid RNA after agarose gel electrophoresis. Panel A: Ethidium bromide stain of agarose gel before transfer. In panel B, hybridization of nucleic acid on DBM paper was carried out with [32 P]-labeled Fragment B (AvaI-BglII) using 10^6 cpm per ml of hybridization solution in 10% dextran sulfate. The DBM paper was then washed at 42° with 2X SSC. This mild washing condition was employed because the expected hybridizable sequence is only 77 nucleotides. Autoradiography was carried out at -20° for seven days. In panel C, hybridization was carried out with [32 P]-labeled 0.9 Kb fragment of Hha I x pOM502. (10^5 cpm per ml of hybridization solution in 10% dextran sulfate.) Washing was carried out under stringent conditions (three times at 68° with 0.5X SSC) to reduce background. The filter containing lanes a and b was exposed to X-ray film at -20° for 24 hours while lane c was autoradiographed for two hours due to the strong intensity for the signal. Lane a contains oviduct poly(A) cytoplasmic RNA (50 μ g), lane b is oviduct poly(A) nuclear RNA (100 μ g) and lane c is 0.9 Kb Hha I x pOM502 DNA fragment (1 μ g).

DNA fragment (Fig. 1) was labeled with [32 P] only at the Hpa II site and was allowed to hybridize to oviduct poly(A) RNA as described in Figure 2. Digestion with either S1 nuclease or Mung bean nuclease was carried out after hybridization. The protected fragments were then subjected to electrophoresis on a sequencing gel alongside a sequence ladder for the coding strand of the Dde I-Hpa II probe (Fig. 4).

After S1 nuclease digestion (Fig. 4, lane S1), the smaller protected fragments migrated as a group at positions -2 to +3 of the DNA sequence shown in

Figure 4: Localization of the 5' ends of oviduct ovomucoid RNA using DNA sequencing gel electrophoresis. The autoradiograph of a 20% sequencing gel of the Hpa I-Dde I DNA fragment is presented. The sequence (coding strand) is read from the 5' end (Hpa II cleavage site) at the bottom of the gel to the 3' end at the top. The SI nuclease digestion product of hybrids between oviduct poly(A) RNA and the Hpa II-Dde I probe was run alongside the ladder sequence (Lane S1) so that the starting nucleotide of each protected fragment can be precisely localized. The Mung bean nuclease digestion product was similarly analyzed (Lane MB).

Fig. 1. The larger fragments also appeared as a group of three bands differing by one or two nucleotides and migrating at positions between -80 and -85. The appearance of several bands differing by one nucleotide also has been observed by others when SI nuclease-generated [32 P] DNA-RNA hybrids were analyzed on high resolution sequencing gels (22,23). Our experiments were carried out under SI nuclease excess conditions, and no change in the banding pattern of the protected fragments was found when we further increased the SI nuclease concentration or prolonged the time of incubation. In interpreting nuclease protection experiments a correction of one-half base has to be made (23). This is because the mobility of the 3'-OH-terminated SI fragments is slightly slower than that of their 3'-P-terminated counterparts produced by chemical cleavage. Thus, the

principal bands of the smaller group of fragments correspond to the nucleotides G and T at positions -1 and +1 on the ovomucoid DNA. In case of the larger fragments, the predominant bands correspond to the nucleotides T and G at positions -84 and -83. Significantly, using yet another DNA probe of differing size (Dde I-Hpa II) the results obtained are in excellent agreement with our previous observations.

Due to the possible ambiguity associated with the S1 nuclease mapping procedure, we decided to use another single-strand specific nuclease, Mung bean nuclease (24) to confirm the results of our RNA mapping studies. The 5'-terminal mapping result with Mung bean nuclease is shown in Figure 4, lane MB. The predominant [³²P]-DNA band protected by chicken oviduct mRNA was found to correspond only to nucleotide T at position +1. A single protected band was also observed at position -84.

Presence of Additional Minor Initiation Regions

After prolonged autoradiography (10 days), a third protected fragment of about 40 nucleotides in length could be observed in Figure 2A, lane a. This indicated the presence of another minor transcription initiation region for chick ovomucoid which is located approximately 20 nucleotides upstream from the major initiation site. This finding was further confirmed by S1 nuclease mapping using [³²P]-labeled Fragment C as probe (Figure 2B, lane a). A faint band of 155 nucleotides was visible after prolonged exposure, and this is again larger than the predominant band (135 nucleotides) by exactly 20 nucleotides. According to densitometer scanning (Figure 2C) we estimate that less than 0.5% of the ovomucoid RNA initiates at this third region. By comparing with the sequence ladder in Figure 4, this third minor region is found to be located at nucleotide -23 of the ovomucoid gene. It is possible that some chicken ovomucoid transcripts can arise from other yet unlocated initiated regions, but their occurrence may be so infrequent that they escape detection because of the sensitivity limits of our S1 nuclease mapping techniques.

Transcription Initiation of Ovomucoid mRNA at the Heterogeneous Regions is Not Differentially Regulated by Specific Steroid Hormone.

Since ovomucoid production in the oviduct is regulated by both estrogen and progesterone (2,3) we wished to investigate whether the synthesis of the heterologous species of ovomucoid mRNA is steroid hormone specific. All previous experiments were carried out using poly(A) mRNA isolated from estrogen-stimulated chick oviduct. In order to map the 5' ends of ovomucoid transcripts in progesterone stimulated oviduct, estrogen stimulated chickens were withdrawn from hormone for 14 days and then restimulated by subcutaneous injections of

Figure 5. S1 nuclease mapping of the 5' ends of ovomucoid transcripts from estrogen- or progesterone-stimulated oviducts [^{32}P]-labeled protected fragments arising from the hybridization of the Hpa II-Taq I probe with 50 μg total mRNA from estrogen-stimulated oviduct (lane a) and with a same amount of total mRNA from two individual progesterone-stimulated oviducts (lanes b and c) were subjected to electrophoresis using a 10% polyacrylamide and 7M urea gel. Lane M is [^{32}P]-labeled Hinf I X ϕ X 174 marker. Autoradiography was carried out at -20° for 12 hrs using a Kodak XAR-5 film.

progesterone (2 mg/day) for 7 days. Oviducts from these animals were removed and poly(A) mRNA were extracted from the tissue as previously described (8,9). S1 nuclease mapping experiments were conducted using mRNA from chick oviduct stimulated by either estrogen or progesterone and the [^{32}P]-labeled Hpa II-Taq I fragment as probe (Fig. 5). Two predominant fragments, of sizes 104 and 20 nucleotides respectively, were similarly detected from both mRNAs as seen also in Fig. 2A, lane a. We thus concluded that the size and distribution of the different species of ovomucoid transcripts are similar when induction is accomplished using either estrogen or progesterone.

DISCUSSION

We have identified at least three species of ovomucoid RNA from chicken oviduct which possess heterogeneous 5' ends. Transcription of about 89% of the ovomucoid RNA appears to initiate around position 1 of the DNA sequence presented in Figure 1. At 30 nucleotides upstream from this position is a heptanucleotide sequence, TATATAT. Such A-T rich regions, commonly known as TATA, or "Goldberg-Hogness" boxes, have been found at positions 21-31 nucleotides upstream from the 5' termini of most eucaryotic genes transcribed by RNA polymerase II (7).

They have been implicated as essential for accurate initiation of transcription (25). A small percentage of the in vivo ovomucoid RNA is initiated at a different region 84 nucleotides further upstream from the major one (position -84, Figure 1). It is perhaps more than a coincidence that beginning at 29 nucleotides upstream (5') from this minor region there is also an A-T rich sequence (TAAATAT) resembling the previous TATA box, but differing by an A to T nucleotide change at the third position. A third region which accounts for initiation of a more infrequent species of ovomucoid RNA was also located.

The fact that the minor ovomucoid RNA species is larger than the dominant species would argue against the possibility of RNA processing being the cause for the existence of the minor transcript. Nevertheless, at the present time we cannot eliminate completely the possibility that all ovomucoid transcripts might be initiated at the minor site(s) and the majority of these molecules are processed rapidly so that they possess a shorter 5' non-coding sequence. However, the observation that the heterogeneous species of ovomucoid RNA are present at the same proportions in both the oviduct nuclear and cytoplasmic RNAs would make this possibility unlikely. Also, recent results from in vitro transcription studies have shown that all detectable transcripts were initiated at position +1 of the ovomucoid gene (S.Y. Tsai, unpublished observation), indicating that this is an authentic initiation region. It has been determined previously that the ovomucoid gene is unique within the chicken genome (5) and so heterogeneous species of ovomucoid RNA cannot be the products of multiple genes. Furthermore, these species may not be due to expression by alleles of the ovomucoid gene because mRNA extracted from the oviducts of three chickens have shown the presence of these species in a similar ratio. Finally, in case of multiple genes or alleles, we would expect the 5' flanking regions to be dissimilar due to a high rate of divergence. Our results also cannot be explained as artifacts due to the folding back of the DNA probe. We used four DNA fragments of different sizes as probes, the smallest one being only 225 (Dde I-Hpa II) bases and obtained identical results. DNA sequence data of this small DNA fragment do not reveal any significant complementary sequence.

Multiple 5' ends have been described for several viral mRNAs. It has been established that the SV40 late messengers have heterogeneous 5' termini encoded in a 400 residue region adjacent to the replication origin (26-28). Heterogeneity has been reported for polyoma late mRNAs (29) and for adenovirus-2 mRNA EII, EIII, EIV, and the protein IX mRNA (30). So far, it is not known

whether such heterogeneous 5' termini are typical among eucaryotic mRNAs. Recently, heterogeneity at the 5' ends of eucaryotic cellular mRNAs has been observed for yeast iso-1-cytochrome c (31) and chicken lysozyme (32). Our preliminary findings from S1 nuclease mapping of chicken ovalbumin mRNAs also reveal the existence of several very minor additional 5' termini, indicating that heterogeneity in the 5' end of ovomucoid RNA is not unique. The limitation in detecting heterogeneous species of a particular eucaryotic mRNA will depend on the concentration of the minor species present within the total population of mRNA. It is conceivable that there are other species of the ovomucoid RNA with still different 5' termini, but they are initiated so infrequently that they cannot be detected by our present methods of analysis.

Recently, a great deal of interest has been channeled towards attempts to understand and define the eucaryotic promoter. It has been proposed that the TATA box is essential for accurate initiation of transcription, and that there may be additional sequences upstream which can modulate transcription. Our findings that transcription initiation in eucaryotic genes is heterogeneous rather than unique further complicates the envisioned picture of the promoter region. It does, however, offer supplementary information that must be considered when the function and structure of the eucaryotic promoter are being defined. Recent studies of the mouse α -amylase gene have revealed that two different promoters can be present for a single cellular gene. Furthermore, the differential use of these promoters is tissue specific (33,34). This finding leads one to consider the possibility that multiple transcription initiation regions could play a part in the regulation of gene expression.

ACKNOWLEDGEMENTS

We thank Kristen Anderson, Sharen Knudsen, and Alan Lidsky for excellent technical assistance. S.L.C.W. is an investigator of the Howard Hughes Medical Institute. This work is supported by NIH grants HD-8188, GM-29697, and the Baylor Center for Population Research and Reproductive Biology.

*Present address: Laboratory of Experimental Pathology, National Cancer Institute, Bethesda, MD 20205, USA

REFERENCES

1. Feeney, R.E. and Allison, R.G. (1969) *Evolutionary Biochemistry of Proteins* (New York, John Wiley)
2. Palmiter, R.D. (1972) *J. Biol. Chem.* 247, 6450-6461
3. Chan, L., Means, A.R. and O'Malley, B.W. (1978) *Vitamins and Hormones* 36, 259-295

4. Hynes, N.E., Groner, B., Sippel, A.E., Nguyen-Huu, M.C. and Schitz, G. (1977) *Cell* 11, 923-932
5. Tsai, S.Y., Roop, D.R., Tsai, M.-J., Stein, J.P., Means, A.R. and O'Malley, B.W. (1978) *Biochemistry* 17, 5773-5780
6. Stein, J.P., Catterall, J.F., Woo, S.L.C., Means, A.R. and O'Malley, B.W. (1978) *Biochemistry* 17, 5763-5772
7. Lai, E.C., Stein, J.P., Catterall, J.F., Woo, S.L.C., Mace, M.L., Means, A.R. and O'Malley, B.W. (1979) *Cell* 18, 829-842
8. Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M.-J. and O'Malley, B.W. (1978) *Cell* 15, 671-685
9. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412
10. Bailey, J.M. and Davidson, N. (1976) *Anal. Biochem.* 70, 75-85
11. Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5350-5354
12. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci., USA* 74, 506-564
13. Sanger, F. and Coulson, A.R. (1977) *FEBS Letters* 87, 107-110
14. Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) *Cell* 21, 681-687
15. Catterall, J.F., Stein, J.P., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) *J. Cell Biol.* 87, 480-487
16. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732
17. Weaver, R.F. and Weissman, C. (1979) *Nucleic Acid Res.* 7, 1175-1195
18. Hentschel, C., Irminger, J.-C., Buckner, P. and Birnstiel, M. (1980) *Nature* 285, 147-151
19. McReynolds, L.A., O'Malley, B.W., Misbet, A.D., Fothergill, J.E., Givole, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) *Nature* 273, 723-728
20. Malek, L.T., Eschenfeldt, W.H., Mums, T.W. and Rhoads, R.E. (1981) *Nuc. Acid Res.* 9, 1657-1673
21. Tsai, M.-J., Ting, A.C., Nordstrom, J.L., Zimmer, W., and O'Malley, B.W. *Cell* 22, 219-230
22. Moss, T. and Birnstiel, M.L. (1979) *Nucl. Acids Res.* 66, 3733-3743
23. Sollner-Webb, B. and Reeder, B. (1979) *Cell* 18, 485-499
24. Dodgson, J.B. and Wells, R.D. (1977) *Biochemistry* 16, 2374-2379
25. Tsai, S.Y., Tsai, M.-J., and O'Malley, B.W. (1981) *Proc. Natl. Acad. Sci. USA* 78, 879-883
26. Groner, Y., Carmi, P. and Aloni, Y. (1977) *Nucleic Acids Res.* 4, 3959-3968
27. Ghosh, P., Reddy, V., Swinscoe, J., Lebowitz, P. and Weissman, S. (1978) *J. Molec. Biol.* 126, 813-846
28. Canaani, D., Kahana, C., Mukamel, A. and Groner, Y. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3078-3082
29. Cowie, A., Tyndall, C., and Kamen, R. (1981) *Nucleic Acid Res.* 9, 6305-6322.
30. Ziff, E.B. (1980) *Nature* 287, 491-499
31. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D., and Smith, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2258-2262
32. Grez, M., Land, H., Giesecke, G., and Schutz, G. (1981) *Cell* 25, 743-752
33. Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) *Cell* 21, 179-187
34. Schibler, V., Tose, M., Pittet, A.-C., Fabiani, L. and Wellauer, P.K. (1980) *J. Mol. Biol.* 142, 93-116