Volume 8 Number 1 1980 Nucleic Acids Research

The ovalbumin gene - sequence of putative control regions

C.Benoist, K.O'Hare, R.Breathnach and P.Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, Strasbourg, France

Received 8 November 1979

ABSTRACT

We present the sequence of regions of the chicken ovalbumin gene believed to be important in the control of initiation of transcription, splicing, and transcription termination or polyadenylation. Comparison with corresponding areas of other genes reveals some homologous regions which might play a role in these processes.

INTRODUCTION

We have previously reported the results of sequence analysis of the split chicken ovalbumin gene showing sequence homologies with other mRNA coding genes at intron-exon junctions (1) and before the presumed start of transcription (2). In this paper we report the results of more extensive analysis at the 5' end of the gene, the sequence of three introns in their entirety and the sequence beyond the 3' end of the ovalbumin mRNA-coding sequences. The roles that these sequences might play in the control of initiation of transcription, splicing of the primary transcript, and termination of transcription and polyadenylation are discussed.

MATERIALS AND METHODS

Ovalbumin genomic DNA fragments were purified as described elsewhere (1, 3). Nucleotide sequence determinations were carried out using the technique of Maxam and Gilbert (4) as modified in Tonegawa et al. (5). Electrophoresis of cleavage products was primarily on 8 % polyacrylamide-8M urea thin gels as described by Sanger and Coulson (6). Sources of materials used have been

described elsewhere (1). Further details on the sequencing procedure and autoradiograms of the sequencing gels are available upon request.

RESULTS AND DISCUSSION.

Sequence at the 5' end of the ovalbumin gene.

Studies on the major late adenovirus 2 transcription unit have led to the conclusion that in this case the start of transcription most probably corresponds to the first, capped nucleotide of the mature messenger (7, 8). Such strong evidence is unfortunately lacking for other genes, including the ovalbumin gene. However, there is no evidence for transcription immediately upstream of the first nucleotides found in mature ov-mRNA, and all nuclear ovalbumin precursor RNA molecules (9) seem to have the same 5' end as ov-mRNA (M. LeMeur, personal communication). Similarly, for the silk fibroin and rabbit β-globin cases, the precursor RNA and the mature mRNA have been reported to have a common 5'-end (10, 11). We assume for the discussion below that the transcription initiation site does correspond to the first nucleotide of the mature messenger.

Extensive biochemical and genetic studies have indicated that in prokaryotes there are two regions of sequence homology just upstream of the transcription initiation site where promoter mutations map (12, 13). At about 10 bp before the initiation site there is an AT-rich region [typically TATAAT (12)], and centred around -35 bp a second sequence, the "recognition site" [typically TGTTGACAATTT (12, 13)]. The sequences upstream from the presumed initiation site for several eukaryotic mRNA coding genes are now known (2, 10, 14 to 20). Hogness and Goldberg (personal communication) have pointed out that around position -20 to -30 there is often an AT-rich region (usually related to the sequence TATAAAA) bounded on both sides by GC-rich sequences. We have previously reported the sequence of the 60 bp before the presumed transcription initiation site of the ovalbumin gene and detected just such a sequence at -25 to -32 bp (2). In a search for additional sequence homologies with other mRNA-coding genes we have extended our sequence determinations at the 5'-end of the ovalbumin gene.

A map of the chicken ovalbumin gene is shown in Fig. 1 (a). The leader coding sequences (L) specify nucleotides 1-47 of the ovalbumin mRNA (1, 2). A 1.7 kb EcoRI-PstI fragment [Pst4-Eco6, see Fig. 1 (a)] containing these sequences has been subcloned into pBR322 from a previously described 16.7 kb fragment containing the entire gene (2) which was isolated from a chicken bank constructed by Dodgson, Strommer and Engel (21). A restriction map of this 1.7 kb fragment is shown in Fig. 1 (b). We have sequenced extensive regions around the leader coding sequences using the Maxam and Gilbert technique (4). (The

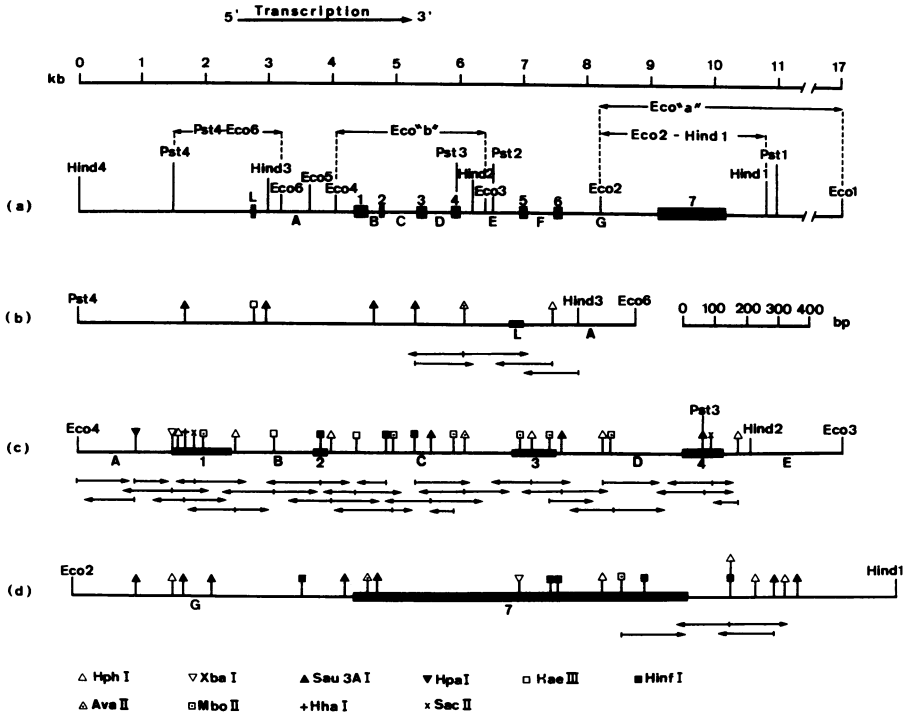

Fig. 1 : a)Map of the chicken ovalbumin gene (1, 2) and scale in kb. The eight exons (L and 1-7) and the seven introns (A to G) are shown. The location within the gene of EcoRI fragment "b", the 1.7 kb Pst4-Eco6 and the 2.6 kb Eco2-Hind1 fragments used for sequencing are indicated.b), c)and d) show restriction maps of the Pst4-Eco6, EcoRI "b" and Eco2-Hind 1 fragments respectively. Arrows indicate the sites used for 5'-end-labelling and the direction and extent of the sequence determinations.

sequencing strategy is outlined in Fig. 1 (b) by the horizontal arrows). In this way we have extended our previous determination to nucleotide number 300 upstream from the 5'-end of the leader-coding sequences and to nucleotide 170 downstream from them (Fig. 2).

Fig. 3 shows parts of the sequences in the chicken ovalbumin, chicken conalbumin (14), adenovirus 2 major late (15) and early 1A (16) genes upstream from the presumed start of transcription. Further upstream than the AT-rich region described above, there are regions of homology between the conalbumin and adenovirus 2 major late genes [as previously pointed out (14)] and between the ovalbumin and early 1A genes. These homologies between pairs of genes are the most extensive we have been able to discern, but by considering all those genes where the 5'-end of the RNA has been defined, a sequence around positions -70 to -80 seems to have been conserved. This is shown in Fig. 4. It must be admitted that the position of the sequence in the conalbumin gene is somewhat displaced relative to the other genes, and that, for instance, the conalbumin and adenovirus late genes share larger sequence homologies at other positions (see Fig. 3). However, both its ubiquity and its similarity to the prokaryotic recognition sequence are noteworthy.

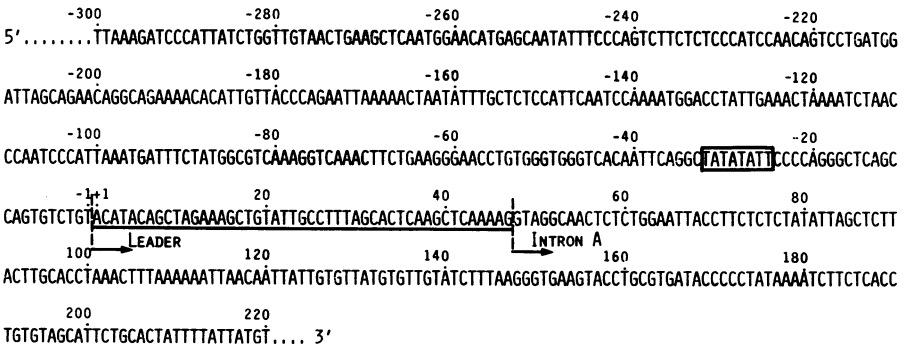

Fig. 2 : Nucleotide sequence around the leader coding region of the ovalbumin gene (Only the non-coding strand is shown). The leader region is underlined and the AT-rich region discussed in the text is boxed. Nucleotides are numbered from the 5' end of the leader region and negatively upstream from this position.

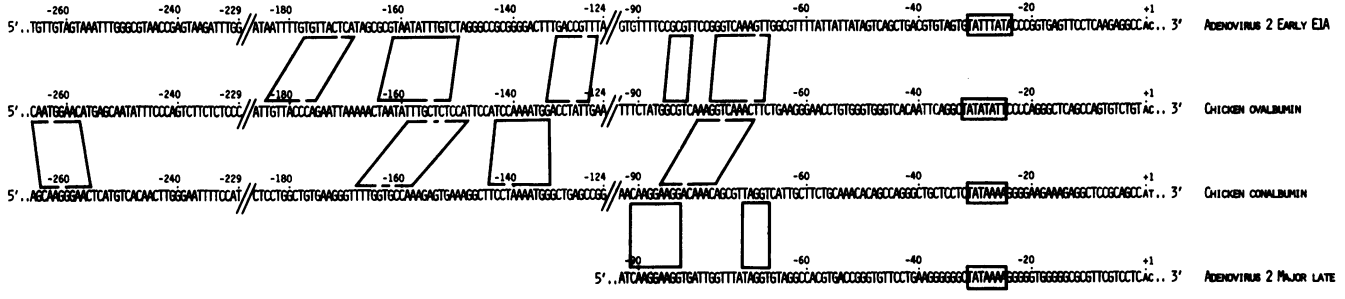

Fig. 3 : Comparison of gene sequences upstream from the AT-rich region (only the non-coding strand is shown). The AT-rich regions are boxed and sequence homologies between pairs of genes are shown by the lines. Gaps in these lines indicate mismatches or deletion/insertions in the homologies. Note that the sequences presented are not contiguous - no strong homologies were detected in the sections of sequences not shown. See text for references.

GGTCAAA$^{-68}_{G}$T	ADENOVIRUS EARLY 1A (16)
GGTCAAA$^{-69}_{A}$CT	CHICKEN OVALBUMIN
GGACAAA$^{-76}_{C}$A	CHICKEN CONALBUMIN (14)
GGTTTATA$^{-69}_{G}$	ADENOVIRUS MAJOR LATE (15)
GGCCAATC$^{-70}_{T}$	RABBIT GLOBIN (17)
GGCCAATC$^{-71}_{T}$	MOUSE GLOBIN (20)
GGCTCATT$^{-70}_{T}$	S.PURPURATUS HISTONE H2B (18, 19)
GTTTATT$^{-68}_{C}$	SILK FIBROIN (10)
5'-GG$^{C}_{T}$CAATCT-3'	MODEL EUKARYOTIC SEQUENCE
5'-TGTTGACAATTT-3'	MODEL PROKARYOTIC RECOGNITION SEQUENCE (13)

Fig. 4 : Sequence homologies around positions -70 to -80 from the presumed site for initiation of transcription. A model eukaryotic sequence that can be derived is shown, along with the prokaryotic recognition sequence. Numbers refer to the position of the last nucleotide in the sequence. Sources of data are as shown.

Although nothing is known of the contacts made by the eukaryotic RNA polymerase B with the DNA during the initiation of transcription, it is certainly possible that interactions are made at positions -80, -30 and 0. Indeed, there is evidence that the smaller E.coli RNA polymerase can interact with DNA over a length of 80 bp during the initiation of transcription (22). The significance, if any, of these sequence homologies may become clearer when the sequence requirements for specific initiation of transcription are known. The availability of *in vitro* transcription systems where specific initiation occurs, such as that recently described for the major late adenovirus 2 transcription unit (23) and the use of reversed genetics (24) to make mutations in putative promoter sequences should ultimately allow definition of the sequence elements important in the initiation of transcription by RNA polymerase B.

The expression of the chicken ovalbumin, conalbumin, ovomucoïd and lysozyme genes in the oviduct can be induced by the steroïd hormones, oestradiol and progesterone (25), although the exact way they respond to induction is different (26). If, by analogy with prokaryotic systems, this control occurs by the interaction of specific factors and RNA polymerase B around the site where transcription initiates, there could be additional sequences shared by these genes. We have compared the sequences of the ovalbumin and conalbumin genes up to position - 300 and found short blocks of conserved sequences around positions -80, -140, -160 and -260 (Fig. 3). However, there is as much sequence homology between the adenovirus 2 early 1A gene and the ovalbumin gene in this region as between the two chicken genes. When additional data is available for the chicken ovomucoïd and lysozyme genes it may become clearer whether these homologies indicate regions important in the hormonal control of oviduct gene expression, although functional studies will be needed to definitively establish any such relationship.

Sequences of three chicken ovalbumin gene introns.

Our previous sequencing studies of the ovalbumin gene exon-intron junctions led us to propose that all exon-intron junctions are related in sequence and may be derived from a

prototype sequence of 5-6 nucleotides, and that the 5'-and 3'-ends of introns are defined by the dinucleotides GT and AG, respectively (1). As sequences related to this prototype sequence occur elsewhere in ovalbumin and other gene exons, it is unlikely to be a sufficient signal to account for splicing specificity. It has been proposed that secondary structure in the primary transcript could bring into close contiguity the ends of exon transcripts to be spliced together (eg. Ref. 27). To see if such a model might apply in the case of the ovalbumin gene we have sequenced in their entirety three of the gene's seven introns.

The EcoRI fragment "b" of the chicken ovalbumin gene has been cloned, mapped and characterised (28). A restriction map of this fragment is shown in Fig. 1 (c), together with the strategy we have used to sequence on the fragment using the technique of Maxam and Gilbert (4). As shown in Fig. 1 (c) we have sequenced more than 75 % of the DNA on both strands and overlapped every restriction site used for sequencing. Areas not sequenced on both strands were invariably checked by repeated sequencing on one strand. The areas corresponding to exons have also been sequenced on a cloned double-stranded cDNA (3).

The sequence we obtained is presented in Fig. 5. The sequence of an independently cloned EcoRI fragment "b" has been reported by Robertson et al. (29). There are 11 differences between the two sequences, at positions 7, 22, 336, 480, 878, 1009, 1260, 1286, 1320, 1533 and 1723. Those at positions 7, 336, and 878 create TaqI, HhaI and HaeIII sites respectively, which are present in our genomic clone but absent from that sequenced by Robertson et al. (29, and our unpublished observations). The remaining differences do not modify any sites. We and others(30-32) have previously provided evidence for allelic variants of the introns of the ovalbumin gene. The differences between our sequence and that of Robertson et al.,(29)(who sequenced less than 55 % of their fragment on both strands), could represent further such variants or sequencing errors, although we are confident of our assignments at these positions. Assuming the differences are not due to error, there are eight nucleotide differences between the two cloned "b" fragments in 1576 bp of

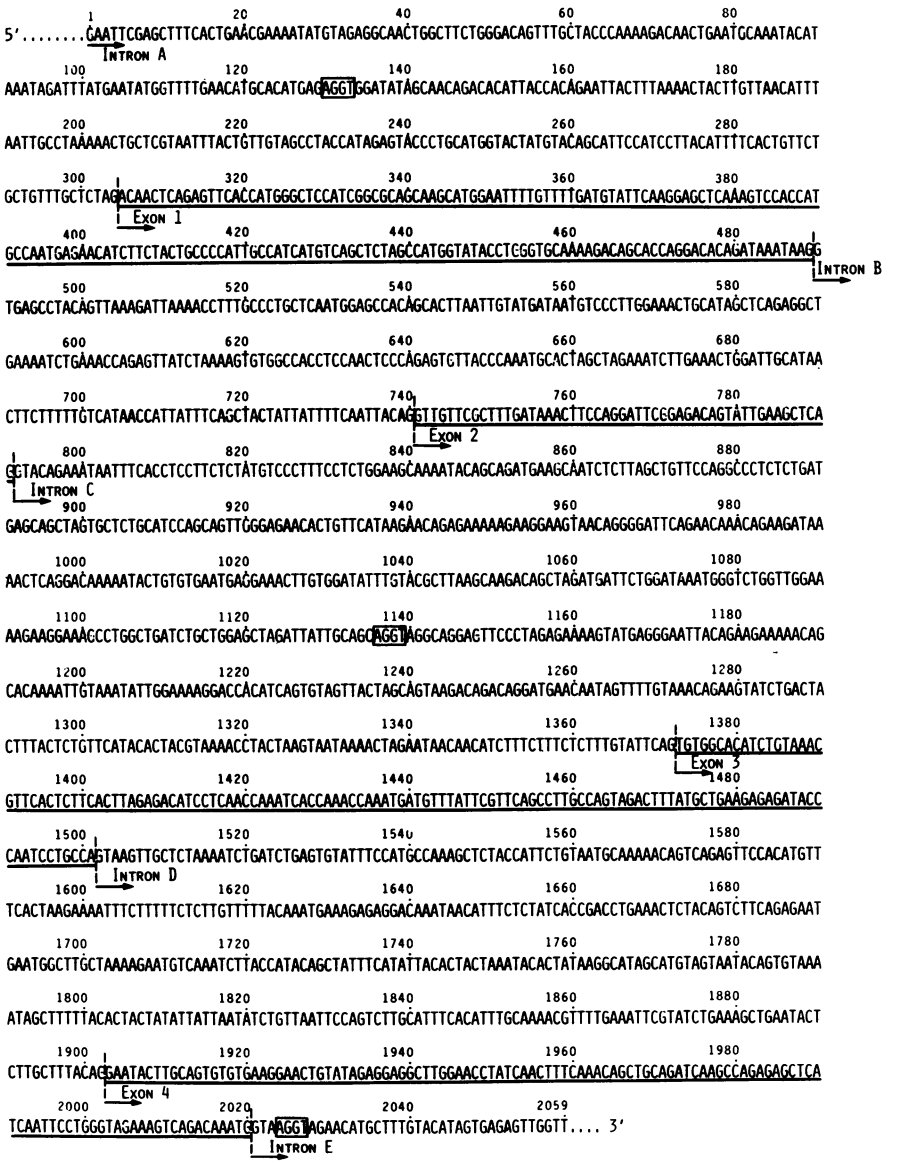

Fig. 5 : Nucleotide sequence of the EcoR1 fragment "b". The sequence (non-coding strand only) shows from the EcoR1 site (Eco4) the 3'-end of intron A, exon 1, intron B, exon 2, intron C, exon 3, intron D, exon 4 and the 5'-end of intron E. The exon sequences are underlined, assuming that the GT-AG rule (1) defines the splice points. The tetranucleotide AGGT discussed in the text is boxed. Nucleotides are numbered from the 5'-end of the EcoR1 recognition sequence.

intron sequenced. We have previously reported a difference rate of six nucleotides in 1872 between ovalbumin mRNAs from two chickens of different breeds (3). Thus the frequency of variation of intron sequences does not appear to be much higher than that found in exons.

Secondary structures.

One of our aims in sequencing the EcoRI fragment "b" was to search for features besides the few conserved nucleotides at splice points which might facilitate splicing. A number of workers have proposed models of secondary structure for individual splicing events in SV40 (33) and adenovirus 2 (15, 34). Using a programm devised and kindly provided by Dr. R. Portmann, we undertook a computer analysis of possible secondary structures in the EcoRI fragment "b" transcript to search for analogies or other structures which might help to align the ends of exon transcript to be spliced together [we consider it admissible to examine the "b" transcript alone as opposed to the entire gene transcripts as recent studies on mouse β-globin-SV40 recombinants have shown that an intact intron and only parts of the two surrounding exons are sufficient for splicing (35)]. Analysis of the computer print-out did not allow us to formulate any more than marginally stable structures which might have facilitated the splicing process. A similar conclusion was reached by Robertson et al. for their sequence (29). This result is perhaps not surprising in the light of recent results on deletion mutants of SV40 which suggest that correct splicing occurs even when large portions of introns have been deleted (36, 37). In addition, using a mouse β-globin-SV40 hybrid, Hamer and Leder have shown that only 13 or fewer nucleotides of exonic sequence are required at the 5'-end of a splice point for correct splicing (35). These results suggest that further computer searches for secondary structures capable of facilitating the splicing process will not be very profitable, at least until the minimal sequences around a splice point necessary for correct splicing have been defined experimentally.

Possible splicing intermediates.

Individual intron transcripts might be removed from precursor RNA molecules in a one step process, or in a discrete number of steps. There is evidence that individual intron transcripts are not necessarily removed in one step in the processing of adenovirus 2 (38), globin (39), and ovalbumin nuclear RNAs (M. LeMeur, personal communication). If the GT-AG rule also applies to splicing for stepwise removal of a given intron transcript it is clear that the sequence AGGU must be present in the transcript at the intermediate splicing point, as this site must serve as both an acceptor (reading AG) and a donor (reading GU) site [terminology of Seif et al. (40)]. We have scanned the sequence of the EcoRI fragment "b" for the tetranucleotide AGGT, and the three cases we have found in introns are boxed in Fig. 5.

The tetranucleotide is not present in introns B and D, but is present in introns A, C and E. The environment of the AGGT in intron C is particularly interesting as there are sequence homologies between it and the environments of the exon 2-intron C and intron C-exon 3 splice points (Fig. 6).

Studies are in progress to determine whether the AGGT in intron C acts as an intermediate splice point, and to determine whether transcripts of introns B and D are removed in one step, as would be required were the GT-AG rule to apply to splicing

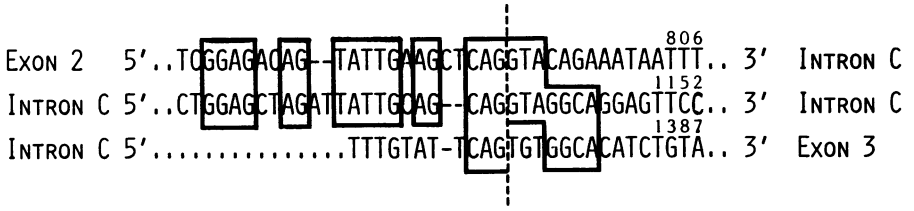

Fig. 6 : A potential internal splice point in intron C. The sequences at the boundaries of intron C are compared with a potential splice point within intron C. The proposed splice points based on the GT-AG rule (1) are shown by the vertical line. The sequences have been aligned to emphasise the homologies (shown boxed) of the putative internal splice sequence with the exon-intron boundary sequences. Dashes show where the homologies are interrupted because there are more nucleotides in one of the sequences than in the others.

intermediates.Clearly, some AGGUs rarely if ever serve as splice points, as the ovalbumin mRNA contains a total of three AGGUs.

Sequences at the 3'-end.

We have previously reported the cloning of EcoRI fragment "a" (41) of the chicken ovalbumin gene (42) [see Fig. 1 (a)]. This fragment was shown by electron microscopy and restriction endonuclease mapping to contain exon 7 of the gene [Fig. 1 (a)] corresponding to the 3'-end of ovalbumin mRNA. We could not, however, determine whether the entirety of the 3'-terminal sequences of ov-mRNA was encoded by this exon. To answer this question and to obtain sequences flanking this exon we subcloned a 2.6 kb EcoRI-HindIII fragment (1) of fragment "a" [see Fig. 1 (a, d)] in pBR 322. This 2.6 kb fragment contains the entirety of exon 7 and some flanking sequences (1). We sequenced the appropriate region of this fragment using the technique of Maxam and Gilbert (4) [see Fig. 1 (d) for details]. The sequence obtained is presented in Fig. 7. By comparison with the known sequence of the ov-mRNA residues proximal to the poly A (43),

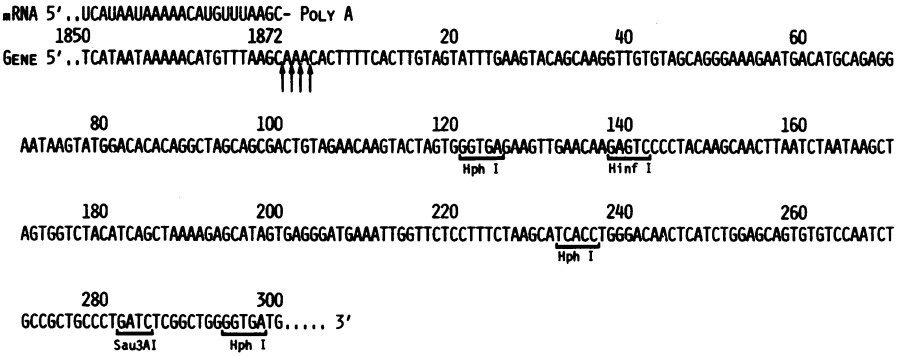

Fig. 7 : Sequence of the chicken ovalbumin gene around its 3'-terminus (the non-coding strand is shown), compared to the 3'-terminal sequence of ovalbumin mRNA. The four vertical arrows indicate the four possible ends of the messenger coding sequence (see text). Nucleotides corresponding to the messenger coding sequence are numbered 1850-1872 according to (3). Nucleotides beyond number 1872 are numbered 1 - 300. Restriction sites which correspond to sites marked on Fig. 1 (d) are indicated below the sequence. The sequence was determined using the technique of Maxam and Gilbert (4) as outlined in the legend to Fig. 1 (d).

it is clear that the entire 3'-terminal sequences of the messenger are encoded by exon 7. It is not possible, however, to define the exact limit of exon 7, as the first one, two or three residues of the poly A tail could be transcribed from the genomic DNA. Four possible ends of exon 7 are therefore indicated in Fig. 7 (vertical arrows). This phenomenon appears to be general : for all cases studied to date the first one [SV40 early mRNA (33) and mouse β-globin major mRNA (20)] or two [rabbit β-globin mRNA (17) SV40 late mRNA (44) and some adenovirus early mRNAs (45, 46)]or three(ovalbumin mRNA) or four [mouse dihydrofolate reductase mRNA (47)] residues of the poly A tail could be transcribed from the genomic DNA (Fig. 8).

Apart from this feature, comparison of the sequences immediately downstream from the last nucleotide coding for messenger for the cases where data is available [SV40 early (33) and late (44)

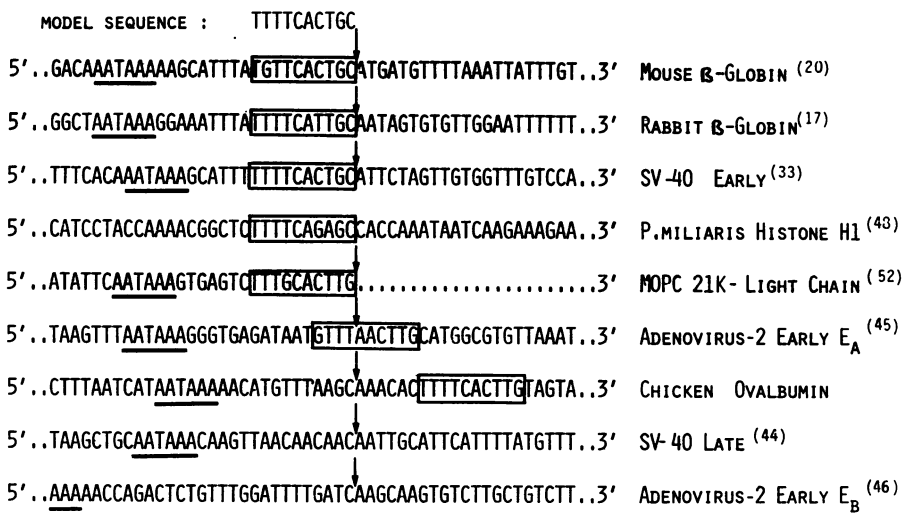

Fig. 8 : Comparison of the 3'-terminal and flanking sequences of various genes. Sequences are aligned such that the first possible non-coding nucleotide lies to the right of the vertical arrows. No arrow is shown for the histone H1 gene, as the precise end of the messenger coding sequence is unknown (48, and M. Birnstiel personal communication). A sequence homology is boxed (see text). The common sequence AAUAAA (54) is underlined (only part of it is shown for adenovirus-2 early E_B mRNA). Sources of data are indicated in the figure.

genes, rabbit (17) and mouse (20) β-globin genes, the ovalbumin gene, sea urchin histone genes (48), two early adenovirus genes (45, 46) and the mouse dihydrofolate reductase gene (47)] shows little sequence homology. These regions are often AT rich, however, and contain some runs of dT residues of varying lengths.

A comparison of sequences at the 3'-end of messengers transcribed by E. coli RNA polymerase has led to a model for transcription termination involving a G-C rich region capable of forming a hairpin, followed by a run of U residues terminating the message (13). These features do not seem to extend to the 3'-ends of eukaryotic mRNAs. Thirteen such sequences are known [human (49, 50) and rabbit (50, 51) α and β-globin mRNA's,mouse β-globin (20) mRNA, two early adenovirus mRNAs (45, 46) a light chain immunoglobulin mRNA (52) ovalbumin mRNA (43) SV40 early (33) and late (44) mRNAs, mouse dihydrofolate reductase mRNA (47) and silk fibroin mRNA (53)]. As previously recognized (54) the only feature common to all 13 mRNAs is the sequence AAUAAA 14-30 residues upstream from the poly A tail. However, the sequence of the 10 nucleotides immediately preceding the poly A tail is clearly related for 5 of these 11 messengers.

This homology is particularly striking for SV40 early mRNAs and mouse β-globin mRNA where 9 of the 10 nucleotides preceding the poly A tail are identical, but it also extends to rabbit and human β-globin messengers, and an immunoglobulin light chain messenger. We illustrate this in Fig. 8 by boxing the appropriate residues of the non-coding strand of the genomic DNA, and by suggesting a model sequence from which the boxed sequences may be derived by at most two changes. All five sea urchin histone genes, which have been recently sequenced (48, and personal communication from M. Birnstiel) contain a sequence related to the model sequence within a few base pairs of the end of the messenger coding sequences. This is shown in Fig. 8 for histone H1 as an example. A sequence related to the model sequence is not found at the 3'-termini of SV40 late mRNAs, two adenovirus early mRNAs, rabbit or human α-globin mRNA, ovalbumin mRNA, silk fibroin mRNA or mouse dihydrofolate reductase mRNA. A related sequence is found in the immediate 3'-flanking region of ovalbumin mRNA and an early adenovirus mRNA

(45), however. The actual sequences found (shown boxed in Fig. 8) are very close to those found at the 3'-end of the immunoglobulin mRNA : 9 out of 10 nucleotides are the same for the ovalbumin case, 7 out of 10 for the adenovirus case. Further studies along the lines of "reversed genetics" (24) will clarify the significance if any of this sequence and of the AAUAAA sequence.

Biohazards associated with the experiments described in this publication were examined previously by the French National Control Committee. The experiments were carried out under L3 B1 conditions (Le Progrès Scientifique, N° 191, Nov. Dec. 1977).

AKNOWLEDGEMENTS

We thank M.C. Gesnel, A. King, J.M. Garnier, E. Taubert and B. Boulay for technical assistance, and Drs M. Birnstiel, U. Petterson, R. Schimke and R. Gelinas for communication of unpublished results. We are particularly grateful to Dr. Rudolf Portmann for helping us to use his computer programs. This work was supported by grants from the INSERM (CRT 76.5.468 and 76.5.462, the CNRS (ATP n° 3558) and the Fondation pour la Recherche Médicale Française. K. O'Hare was supported by an Embo long term fellowship.

REFERENCES

1. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc. Natl. Acad. Sci. 75, 4853-4857.
2. 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.
3. O'Hare, K., Breathnach, R., Benoist, C., and Chambon, P. (1979) Nucl. Acid. Res. 7, 321-334.
4. Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. 74, 560-564.
5. Tonegawa, S., Maxam, A., Tizard, R., Bernard, O., and Gilbert, W. (1978) Proc. Natl. Acad. Sci. 75, 1485-1489.
6. Sanger, F., and Coulson, A. (1977) FEBS Lett. 87, 107-110.
7. Ziff, E., and Evans, R. (1978) Cell 15, 1463-1475.
8. Manley, J., Sharp, P., and Gefter, M. (1979) Proc. Natl. Acad. Sci. 76, 160-164.
9. Chambon, P., Benoist, C., Breathnach, R., Cochet, M., Gannon, F., Gerlinger, P., Krust, A., LeMeur, M., LePennec, J.P., Mandel, J.L., O'Hare, K., and Perrin, F. (1979) in Proceedings of the 11th Miami Winter Symposium "From Gene to Protein : Transfer in Normal and Abnormal cells".

Academic Press, New York.
10. Tsujimoto, Y., and Suzuki, Y. (1979) Cell 16, 425-436.
11. Weaver, R., and Weissman, C. (1979) Nucleic Acids Res., in press.
12. Rosenberg, M., and Court, D. (1980) Ann. Rev. Genetics, in press.
13. Gilbert, W. (1976) in RNA Polymerase, eds. R. Losick and M. Chamberlin, pp 193-205, Cold Spring Harbor Laboratory, N.Y.
14. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P. Nature, in press.
15. Akusjärvi, G., and Pettersson, U. J. Mol. Biol., in press.
16. Maat, J., and Van Ormondt, H. (1979) Gene 6, 75-90.
17. Weissmann, C., Mantei, N., Boll, W., Weaver, R., Wilkie, N., Clements, B., Taniguchi, T., Van Ooyen, A., Van Den Berg, J., Fried, M., and Murray, K. (1979) in Proceedings of the 11th Miami Winter Symposium "From Gene to Protein : Transfer in normal and abnormal cells". Academic Press, New York.
18. Levy, S., Sures, I., and Kedes, L. (1979) Nature 279, 737-739.
19. Sures, I., Lowry, J., and Kedes, L. (1978) Cell 15, 1033-1044.
20. Konkel, D., Tilghman, S., and Leder, P. (1978) Cell 15, 1125-1132.
21. Dodgson, J., Strommer, J., and Engel, J. (1979) Cell 16, 879-887.
22. Schmitz, A., and Galas, K. (1979) Nucleic Acids Res. 6, 111-137.
23. Weil, P., Luse, O., Segall, J., and Roeder, R., Cell, in press.
24. Weissmann, C. (1978) TIBS 3,N109-111.
25. Palmiter, R. (1972) J. Biol. Chem. 247, 6450-6461.
26. Lee, D., McKnight, G., and Palmiter, R. (1978) J. Biol. Chem. 253, 3494-3503.
27. Klessig, D. (1977) Cell 12, 9-21.
28. Garapin, A., Cami, B., Roskam, W., Kourilsky, P., LePennec, J.P., Perrin, F., Gerlinger, P., Cochet, M., and Chambon, P. (1978) Cell 14, 629-639.
29. Robertson, M., Staden, R., Tanaka, Y., Catterall, J., O'Malley, B., and Brownlee, G. (1979) Nature 278, 370-372.
30. Mandel, J.L., Breathnach, R., Gerlinger, P., LeMeur, M., Gannon, F., and Chambon, P. (1978) Cell 14, 641-653.
31. Weinstock, R., Sweet, R., Weiss, M., Cedar, H., and Axel, R. (1978) Proc. Natl. Acad. Sci. 75, 1299-1303.
32. Lai, E., Woo, S., Dugaiczyk, A., and O'Malley, B. (1979) Cell 16, 201-211.
33. Reddy, V., Ghosh, P., Lebowitz, P., Piatak, M., and Weissman, S. (1979) J. Virol. 30, 279-296.
34. Zain, S., Gingeras, T., Bullock, P., Wong, G., and Gelinas, R., J. Mol. Biol., in press.
35. Hamer, D., and Leder, P. (1979) Cell 17, 737-747.
36. Thimmappaya, B., and Shenk, T. (1979) J. Virol. 30, 668-673.
37. Volckaert, G., Feunteun, J., Crawford, L., Berg, P., and Fiers, W. (1979) J. Virol. 30, 674-682.
38. Chow, L., and Broker, T. (1978) Cell 15, 497-510.
39. Kinniburgh, A., and Ross, J. (1979) Cell 17, 915-921.

40. Seif, I., Khoury, G., and Dhar, R. (1979) Nucleic Acids Res. 6, 3387-3398.
41. Breathnach, R., Mandel, J.L., and Chambon, P. (1977) Nature, 270, 314-319.
42. Garapin, A.C., LePennec, J.P., Roskam, W., Perrin, F., Cami, B., Krust, A., Breathnach, R., Chambon, P., and Kourilsky, P. (1978) Nature 273, 349-354.
43. Cheng, C.C., Brownlee, G., Carey, N., Boch, M., Gillam, S., and Smith, M. (1976) J. Molec. Biol. 107, 527-547.
44. Reddy, V., Ghosh, P., Lebowitz, P., and Weissman, S. (1978) Nucleic Acids Res. 5, 4195-4211.
45. Perricaudet, N., Akusjärvi, G., Virtanen, A., and Petterson, U., Nature, submitted.
46. Aleström, P., Akusjärvi, G., Perricaudet, M., Matthews, M., and Petterson, U., J. Mol. Biol., submitted.
47. Nunberg, J., Kaufman, R., Chang, A., Cohen, S., and Schimke, R., in press.
48. Busslinger, P., Portman, R., and Birnstiel, M. (1979) Nucleic Acids Res. 6, 2997-3008.
49. Wilson, J.J., De Riel, J., Forget, B., Marotta, C., and Weissman, S. (1977) Nucleic Acids Res. 4, 2353-2367.
50. Proudfoot, N.J. (1977) Cell 10, 559-570.
51. Heindell, H.C., Lin, A., Paddock, G., Studnicka, G., and Salser, W. (1978) Cell 15, 43-54.
52. Bernard, O., Jackson, J., Cory, C., and Adams, J. (1977) Biochemistry 16, 4117-4125
53. Hagenbuchle, O., Knikeles, M., and Sprague, K. (1979) J. Biol. Chem. 254, 7157-7162.
54. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.