To the Graduate Council:

I am submitting herewith a dissertation written by Harry Ratrie III entitled "Expression of the Albumin and Alphafetoprotein Genes in the Mouse: Analysis of Post-Transcriptional Nuclear Events." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

John Papaconstantinou
Major Professor

We have read this dissertation
and recommend its acceptance:

Robert K Juyamura

Sankar Mitra

Francis T. Kenney

Wan-Kuang Yee

Accepted for the Council:

Carl Minkel
The Graduate School

EXPRESSION OF THE ALBUMIN AND ALPHAFETOPROTEIN GENES IN THE MOUSE: ANALYSIS OF POST-TRANSCRIPTIONAL NUCLEAR EVENTS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Harry Ratrie III

March 1984

ABSTRACT

The expression of the albumin and alphafetoprotein (AFP) genes of the mouse has been studied with particular emphasis on nuclear events following transcription. Using RNA blot and hybridization techniques, high molecular weight nuclear RNA has been examined from three different tissue sources. Nuclear RNA from adult liver was shown to contain multiple albumin specific species that were large molecular weight, polyadenylated and demonstrated a reproducible set of band intensities. When newborn mouse liver and hepatoma tissues, which also synthesize AFP, were analyzed qualitatively similar AFP precursor bands were detected including one that possibly represents the AFP primary transcript. The albumin RNA precursors, when analyzed in tissues that also synthesize AFP, reveal an alteration of the characteristic adult bands that may be related to the expression of a fetal-like phenotype.

Structural analysis of the albumin gene revealed that intron 12 contains a sequence that is highly repeated in the mouse genome and most likely has appeared in this location as a result of an evolutionarily recent transposition event. The intron 12 region and three others that were shown to contain unique sequence regions were subcloned into the plasmid vector pBR322 for use in further studies.

Hybridization of intron regions to nuclear RNA blots suggested that processing of the albumin RNA precursors occurs via a preferred

pathway in agreement with other high abundance mRNA genes. Some of the albumin nuclear RNA precursors were shown to be associated with the nuclear protein matrix, a result which supports a function for this structure as the site for RNA processing. These data are discussed in light of possible regulatory events affecting these genes during RNA processing in the nucleus.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Since the recognition in the early nineteenth century of the cellular basis of life, studies of the lives and natural histories of cells have continued unabated. Although the cell nucleus was early recognized as the largest of the cellular organelles, characterization of the nuclear material lagged behind studies of the structure and function of cytoplasmic components. For example, the fine structure of genes whose nuclear residence and protein products have been recognized for many years are discoveries that have only been made in the past ten years. Modern biochemistry now has resolved various levels of nuclear function, some examples of which are replication of the DNA, assembly of DNA into chromatin in particular transcriptional states, transcription itself, and modification and processing of heterogeneous nuclear RNA (hnRNA). However, most of these processes are still only poorly understood. These events might be considered as discrete nuclear processes that come into focus as researchers adjust their scrutiny up and down through these functional levels. Thus, just as the cytoplasmic events involved in protein synthesis and secretion had been in sharpest focus, now features of the nuclear mechanism for transport of genetic information from the DNA to the cytoplasm are taking their place at the forefront of research in molecular biology.

Early work had established that the genetic information, encoded by DNA, was transcribed into an RNA molecule which was subsequently transported to the cytoplasm. Once there, it could be assembled into a complex with ribosomes and other factors for translation into protein. Measurements of the amount of DNA per nucleus show that it is in great excess of the amount required to code for cellular protein (181). Similarly, the nuclear endowment of RNA, in terms of sequence complexity and size, was shown experimentally to vastly surpass that found in the cytoplasm (47,49). A partial answer to these paradoxes was provided when in 1977 the following observations were reported. First, sections of an adenovirus late messenger RNA (mRNA) sequence was shown by R-loop studies to originate from different regions of the genome (21). Also the gene that codes for chicken ovalbumin was shown to be interrupted by non-coding DNA sequences (29). These observations of split genes were soon confirmed for many other genes (27,28,83). This arrangement of expressed (exon) and intervening (intron) sequences proved to be characteristic of many transfer RNA (tRNA) genes (121), ribosomal RNA (rRNA) genes (57) and most higher eukaryotic genes that code for protein (184). Transcription of these genes produces a primary precursor RNA that contains all the expressed and intervening sequences and these RNAs have been shown to be processed by splicing out the intervening sequences to produce a smaller mature mRNA (28). Genes which do not seem to employ this split organization are those coding for histones (28) and interferons (152). Lower eukaryotes,

for instance yeast, also vary from this organization in having split genes for their ribosomal proteins but very few others (141).

Speculation about the origins of this mosaic gene organization originally revolved around a long standing question of whether eukaryotes or prokaryotes were more highly evolved (145). As new observations were made, the weight of the argument has swung recently toward the view that eukaryotic genomes represent a more primitive type of organization and that prokaryotes probably have eliminated intervening regions as they have become more specialized (51,60). Strong support for this concept came as exon regions were shown to represent functional units for several proteins (7,26,44) and were proposed to be evolutionarily assembled by rearrangements of functional exon groups into a selectable conformation (60,99). While this concept is intellectually attractive, the fifty small exons of the α-collagen gene whose separate individual functions are difficult to imagine, must also be accounted for (122). The molecular mechanisms required to achieve the evolutionary changes of casting off or acquiring intron DNA are still unknown and discussion of the possible origins and functions continue.

For some time prior to the discovery of split genes, a precursor role has been suggested for large molecular weight nuclear RNA (48) and several later reports confirmed this notion experimentally using pulse labelling of nuclear RNA and hybridization with complementary DNA (cDNA) probes (9,45,69,156). Since it had been shown that the nuclear precursor molecules were much larger

than the mature messenger RNA (mRNA) found in the cytoplasm and the intervening sequences are not present in the mRNA, it became obvious that this nuclear transcript is processsed by excising the introns and splicing the exons together. Using RNA blot techniques, intermediates of this splicing process have been detected for several systems (105,119,138,153,158).

How the splicing machinery recognizes the sequences that are to be excised is still a question under investigation. One effort to address this problem was reported by Chu and Sharp (38), who created a gene with a chimeric intron from SV40 and mouse β-globin sequences and found that the transcripts from this gene were correctly spliced in monkey cells. The conclusion from this type of study, at least for higher eukaryotic mRNAs, is that the internal sequences of introns are not important for correct splicing. In contrast, the sequences immediately surrounding the intron/exon junctions do seem to be important as interpreted from their almost total conservation in junctions examined so far. The 5'-dinucleotide of introns is GT while the 3'-dinucleotide is AG and a fairly well conserved consensus sequence of about eight nucleotides surrounds these sequences (28,169). In addition, studies that analyzed a class of β-thalassemias with splice junction lesions showed that cryptic splice sequences that resembled the normal sequence could be used by the splicing machinery (193). A second finding of these studies was that normally inactive sequences that were mutated to resemble legitimate junctions could also be spliced (61). It has

been shown by several groups that there are small RNAs in the nucleus (192) and that these small nuclear RNAs (snRNA), usually found in small nuclear ribonucleoprotein (snRNP) particles, share significant homology with the sequences surrounding exon/intron junctions. As a result of these studies, it has been proposed that this RNA functions as a guide to correctly align the exon/intron junction and form a double stranded complex that can then be accurately spliced (97,124, 152). Supporting data for this model come from experiments showing that antibodies that precipitate snRNPs can inhibit splicing of adenovirus transcripts in vitro (195). Other general models invoking a function for the poly[A] tail (22) or a 5' to 3' scanning function for the splicing enzymes (92) have been contradicted by more recent studies. Darnell's group found that adenovirus transcripts are processed even without poly[A] (199) and the recent finding of cryptic splice sites in globin introns seems to rule out simple scanning models (112) of splice site selection. Other exceptions to the rules observed in higher eukaryotes have been found in simpler organisms. In yeast mitochondria, it has been shown that a partially intron coded maturase is needed to order to properly splice precursors for the cytochrome b mRNA (50). The requirement for an internal sequence of yeast nuclear introns to interact with an snRNA, in order for processing of the precursor RNA for ribosomal proteins, has already been recently described (93,141). Lack of this seven nucleotide (nt) internal sequence is proposed to explain the observation that yeast cannot splice precursors from higher eukaryotes (94).

In Tetrahymena an intron in rRNA seems able to autocatalyze its own removal without involvement of protein (198). The obvious disparity in mechanisms used to process RNA by different organisms and for different classes of RNA in the same organism has led to the proposal that there are three classes of splicing reactions that use different mechanisms and for which different factors are important (35). For transcripts of nuclear genes that code for protein, important factors in the recognition of splice junctions seem to be sequences surrounding the junctions, and the interactions of snRNPs with hnRNPs. Mechanistic studies of RNA processing in higher eukaryotes are now being performed using in vitro or cell-free processing systems capable of systematically analyzing the multiple factors involved in this process (65,126).

As might be expected, the mechanisms that remove introns from RNA precursors serve more than simply a processing role in some cases. RNA precursors for immunoglobulin heavy chains use alternate splice sites that regulate expression by determining whether a membrane bound or secreted form of the protein will be translated (5,54). In the rat, it has been found that in the thyroid gland characteristic processing of a precursor RNA produces calcitonin mRNA, but an alternate processing pathway used by the pituitary on the same primary RNA yields a neuropeptide (called calcitonin gene-related peptide) mRNA (6). Few other instances of regulation by processing have been observed to date, but it has been predicted that more will be discovered as genes that produce low abundance mRNAs are examined (72).

One other partial explanation for the occurrence of more than the required amounts of DNA in eukaryotic cells is the presence of families of repeated sequences. These sequences make up a significant portion of mammalian DNA, e.g., 20-30% in humans (164), and many of these repeats are found distributed throughout the genome, interspersed with single copy sequences (49). Various classes of interspersed repeated sequences may be repeated from a few to many thousands of copies per haploid genome. The length of these repeats is characteristically less than 500 base pairs (bp) and therefore they have been labelled short interspersed repeats (SINES) (172). Their distribution has not as yet revealed any discernible pattern and, therefore, functional assignments have remained theoretical (172). Sequence analysis of several of these families has revealed structural features that are found in many procaryotic transposable elements. The best studied mamalian example is the Alu family of repeated sequences which are 130 bp long in mouse and 300 bp long in humans (164). These sequences possess short flanking direct repeats and poly[dA] sequences and similarities between these structures and procaryotic transposable elements (89) have led to the proposal that these also represent mobile genetic elements (115). General models have been proposed for eukaryotic repeats but those systems with enough data on which to base these theories are very few, and may represent special classes of regulatory elements. For maize (104), Drosophila (23,157), and sea urchin (49) systems, good data supporting phenotypic effects caused by

transposable repeated regions exist, but the underlying molecular rationale, if any (52,125), remains obscure.

For nuclear processes, another level of focus which has developed gradually in recent years, is the description and function of the structural framework of the nucleus. This foundation of nuclear structure, known as the protein nuclear matrix, is composed of a limited set of proteins and is resistant to nuclease digestions and extraction with detergents and high concentrations of salts (17, 20,34). Digestion with nucleases does not destroy the structural integrity of this nuclear framework suggesting that nucleic acids are not a structural component. These treatments leave residual, spherical structures which are visualized with the electron microscope as a fibrous anastomosing network that connects with the surrounding lamina and runs throughout the nuclear interior (19,20). Analogous framework structures also occur associated with metaphase chromosomes and have been called chromosome scaffolds (4,132). This type of structure has now been found in many tissues from many organisms (73,75,102,106). It has been proposed that nuclear DNA may be associated with the matrix in two important capacities. First, a role has been proposed in the binding of DNA at the bases of long, supercoiled loops (15,42,187) having an average length which corresponds quite well to that of replicons for their respective tissues (197). In addition, it has now been reported that newly replicated DNA is associated with the matrix (18,130,187). These studies have led to the proposal that replication of the DNA

occurs at the bases of the observed loops and that the loops are reeled through the matrix-fixed replication complexes as they are copied (16,107,130,187). The second major function described for nuclear matrix is one of an association with the DNA of actively transcribing genes. As increasing amounts of the DNA of nuclear matrices is digested away by DNase treatment, that which remains attached to the matrix becomes more enriched for gene sequences that were active in the tissue of origin (43,114,129,149,150). In addition, selected repetitive sequences are also enriched (174) in these preparations. Some studies present data conflicting with those that show the enrichment of specific DNA sequences (26,11), but these results can be accounted for by differences in preparation (84,174) of the matrix. Newly synthesized nuclear RNA has also been found to be associated with nuclear matrix (75). This observation, together with the often observed association of high molecular weight RNA with this structure (53,73,74,80,109,186), has recently culminated in the proposal that the processing of precursor RNAs also occurs at sites on the matrix (39,103). A unifying concept can be envisioned here if one considers that the multiple and complicated reactions that are known to occur in the nucleus, occur not in a disorganized soup of nucleoplasm but instead, on an organized structural framework. Thus, the matrix model of nuclear function predicts the association of the appropriate gene sequences, whether for replication or transcription, and in the case of transcription, the passing on of the RNA products to the sites for processing.

Products of processing may then be further passed on to the sites for nuclear membrane transport (10). It is recognized that the procedures involved in the preparation of nuclear matrix and the analysis of matrix properties are ones that might easily lead to erroneous conclusions. Because of this fact studies of functions associated with nuclear matrix remain highly controversial. At the present time techniques are not available to unequivocally determine the facts of nuclear matrix function but in at least two recent reports it is felt that the carefully designed controls that were presented strongly supported the authors' interpretations (39, 149).

The studies presented here have focused mainly on the nuclear events involving precursor RNAs for albumin and alphafetoprotein (AFP). These two proteins have long served as useful models for the mechanisms of regulation of protein synthesis and secretion (31, 85,140,178,180) and the expression of oncofetal proteins during carcinogenesis (2,13,168). More recently, structural information for both the proteins themselves (30,139) and the genetic locus from which they are expressed (62,63,88,95) has revealed interesting evolutionary relationships. In rodents, albumin synthesis begins in the fetal liver and increases to its adult level (10% of the total protein synthesis) at the time of birth (140). In the adult, this level of synthesis and secretion maintains albumin as the major serum protein. AFP is expressed very early in the yolk sac of the developing fetus (37) and later in the fetal liver until shortly

after birth when its synthesis falls to undetectable levels (100, 167,183). Hepatocarcinogens (1), hepatotoxins (8,191), and liver regeneration induced by surgical treatment (190) have been shown to induce re-expression of the AFP gene.

CHAPTER II

ANALYSIS OF mRNA PROCESSING INTERMEDIATES FOR ALBUMIN AND ALPHAFETOPROTEIN

I. INTRODUCTION

Most of the genes of mammalian viruses and the structural proteins of higher eukaryotes have now been shown to contain intervening or intron regions of noncoding DNA (28). For various known split genes that code for protein the numbers of introns range from two (86) to fifty (122) and span a length of DNA sequence from three- to tenfold greater than that found in the mature mRNA. Large hnRNA had been known to contain the sequences found in the cytoplasmic mRNA population (98) and the discovery of split genes confirmed, in general, the proposal for the generation of smaller, mature mRNA by an RNA processing mechanism (48). As a first step in the understanding of this process, a number of studies have been performed aimed at characterization of the modification of specific nuclear RNAs. These modifications include the addition of a 5' end cap structure (170,201), methylation of a few selected bases (146), and addition of 3'-end poly[A] tract (3). This poly[A] sequence has been shown to be added very early after transcription (117,163), and its addition seems to be related to a polyadenylation signal sequence 5'-AATAAA-3' found in the 3' region of split genes (143). This sequence is reported to specify the endonucleolytic cutting of the primary

transcript (110) approximately 20 nt downstream (towards the 3' end), at which site the poly[A] tail is added (143). Some non-polyadenylated RNAs have never possessed a poly[A] tail but some RNAs probably have gradually lost poly[A] sequences as they age (108,162). Early polyadenylation suggests a role in subsequent processing reactions (23) but, at least for adenovirus, mRNA precursors have been shown to process efficiently without poly[A] (199).

The remaining major modification of primary transcripts, i.e., the removal of intervening sequences, has also been examined in a number of systems. In a mammalian DNA tumor virus (adenovirus), primary precursors for late stages of transcription are processed through a series of alternate pathways that yield a large repertoire of individual mRNAs from a single RNA precursor (200). For these virus systems, selection of the correct polyadenylation and splice sites, to ensure the proper production of translatable mRNAs, represents a complex regulatory mechanism. SV40, a DNA virus, has also been exploited in proving that splicing not only occurs but may be required. When intron-free SV40 mutants, which were constructed to contain all of the information required for protein expression were transfected into monkey cells, no stable RNA was detected (48) even though transcription occurred (66).

For a number of mammalian protein-coding genes, RNA intermediates of processing have been detected using techniques of hybridizing cDNA probes to nuclear RNAs blotted onto filter paper (105,119,138,153,163). These studies which characterize the

different sizes and patterns of RNA precursors of individual genes have confirmed several expected features of precursor RNAs. First, most intermediates are polyadenylated and exist in discrete reproducible size patterns that, in many of these systems, include a size expected for the primary RNA transcript. Changes in the observed amounts of individual precursors for ovomucoid mRNA were consistent with transcriptional regulation of this RNA during induction by diethylstilbestrol (119). Also, in the immunoglobulin system where DNA recombination events are required to produce the correct arrangement for light chain genes, sizes of the RNA precursor intermediates are indicative of the chosen J region DNA (138). From these studies it has become clear that the number of intermediates seen in the nucleus is not the same as the number of introns and, that in some cases, the removal of individual introns may occur in multiple steps (87). In some cases RNA processing can function in the regulation of gene expression by differentially splicing precursor RNAs to yield more than one type of functional mRNA (5,54). Further details on the order of processing and possible sites of these reactions are discussed in Chapter IV.

Albumin and AFP are serum proteins that are proposed to have evolved from a gene duplication about 300-500 million years ago (79). These genes, closely linked on chromosome five in the mouse, are each composed of fifteen analogous exon regions and fourteen intervening sequences and are separated by a 13.5 kb intergenic region (79) (Figure II-1). In the mouse, the AFP pattern of

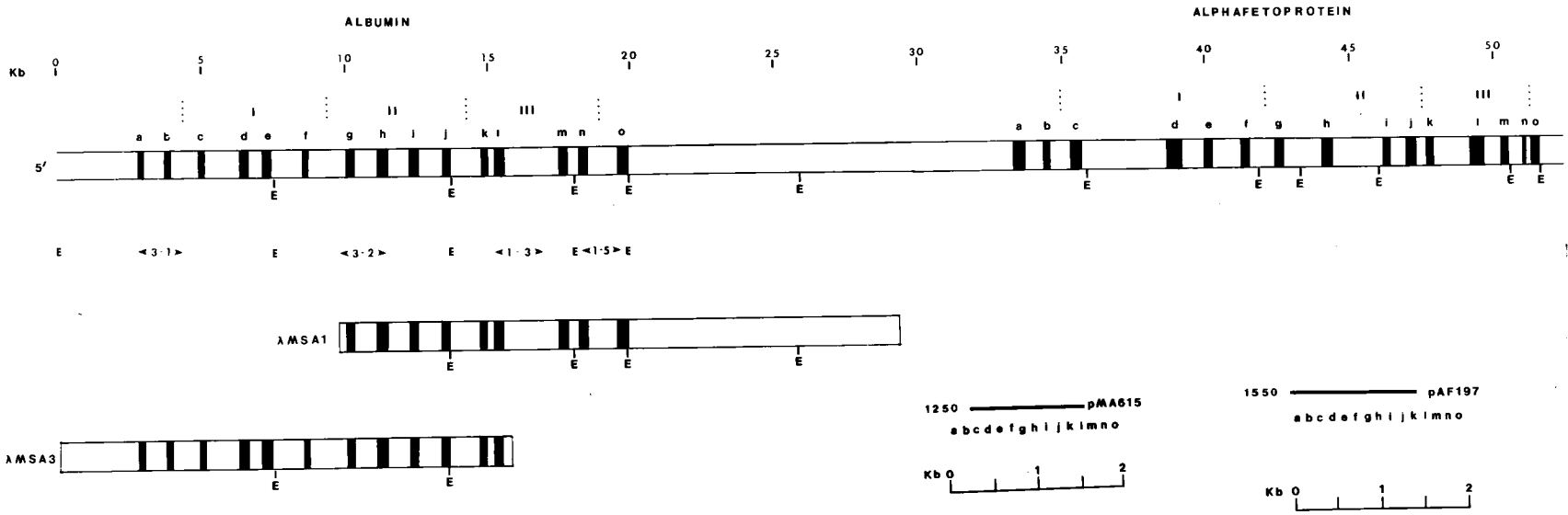


Figure II-1. Map of the albumin and alphafetoprotein gene region.

Scale in kilobases (kb) pairs is given starting from the 5'-end of the MSA3 phage λ clone. Exon regions of the genes are labelled a through o above the genes and Eco RI sites (E) are indicated below the genes. Groups of four exons that comprise the three primordial domains are indicated above the gene map. The two regions of the gene contained in the λMSA1 and λMSA3 clones are shown below the albumin gene. The ends of the MSA clones are Eco RI sites that result from the original cloning and are not contained in the natural gene. The line labelled pMA615 indicates that this 1250 base pair (bp) albumin cDNA clone contains sequences covering exons c through l. The line labelled pAF197 indicates that this 1500 bp alphafetoprotein cDNA clone covers exons a through k. The 4 Eco RI fragments that were subcloned into pBR325 are labelled immediately below the albumin gene.

expression varies from expression in early fetus by the yolk sac (37) and later the liver, to a peak at around birth, then falling to barely detectable levels within several weeks after birth (183). AFP expression can recur after various forms of liver damage (14,100, 167,183) and in association with oncogenic events in the liver (177).

cDNA clones representing most of the mRNA sequences for these two proteins have been employed in studies designed to test the presence of RNA processing intermediates in the nuclei of both fetal and adult liver and in a transplantable minimal deviation hepatoma. In light of the multiple exon structures of these genes, a complex pattern of RNA processing intermediates would be expected. Transfer of nuclear RNA that had been resolved in denaturing agarose gels and hybridization with ^{32}P-labelled cDNA probes has resulted in detection of processing intermediates for both albumin and AFP. We have been able to demonstrate the specificity of these respective probes while also noting similar yet unique bands of large RNA intermediates.

II. EXPERIMENTAL PROCEDURES

Purification of Nuclear RNA

Sterile RNase-free water was prepared by autoclaving a 0.025% solution of diethylpyrocarbonate (DEPC, Sigma Chemical Company) and then boiling the solution down to about 80% of the starting volume to eliminate residual DEPC. All solutions for use in RNA studies were made with this water and then autoclaved if possible. All

glassware and plasticware were rinsed in a 0.1% DEPC solution and then baked at 170°C or 50°C respectively. Solid CsCl was stored in the 170°C oven until used.

Nuclei were isolated from Balb/c mouse livers by preparing a 10% homogenate (w/v) in a solution containing 100 mM NaCl, 50 mM KCl, 20 mM Tris-HCl, 10 mM ethylenediamine tetraacetic acid (EDTA), 0.5 mM phenylmethylsulfonyl fluoride, 0.25 M sucrose and 1.0% nonidet-P40 (NP-40), pH 7.4, but without methylmethane thiosulfonate (149). Nuclei were isolated by pelleting through a cushion of 0.5 M sucrose in the same buffer without NP-40 (149). Pelleted nuclei were immediately lysed using freshly prepared lysing buffer containing 0.35 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, 7 M urea, and 2% sodium dodecylsulfate (SDS), pH 8.0 (155). Samples were dispersed with 8-10 strokes in a Dounce homogenizer. Alternatively, nuclei were prepared using the citric acid procedure of Busch and Daskal (33), using 5% citric acid in the first homogenate. The second nuclear pellet was lysed as above except that only two or three strokes of the Dounce homogenizer were used. The nuclear lysate samples were extracted with phenol, chloroform and isoamyl alcohol (128) and the final aqueous phase recovered. To each milliliter of aqueous phase, 0.6 g of solid CsCl was added and then 3.5 ml of this solution was layered over a 2 ml cushion of 5.6 M CsCl in 0.1 M EDTA. RNA was pelleted by centrifugation at 100,000 g for 16 hours at 20°C in an SW50.1 rotor. Pellets were resuspended in 10 mM Tris-HCl, pH 7.4, 1 mM EDTA (TE), and quantitated at 260 nm using a Beckman UV5260 Spectrophotometer.

Samples were made 0.2 M in sodium acetate and precipitated with two volumes of absolute ethanol at -90°C for sixty minutes or -20°C overnight.

Oligo[dT]-Cellulose Chromatography

RNA samples were selected for poly[A]-containing and poly[A]-lacking sequences (177). The poly[A+] RNA samples were prepared by passage, two times, through an oligo[dT]-cellulose column. The poly[A-] samples were passed over the column three times in order to remove all poly[A+] RNA.

Gel Analysis of RNA

RNA samples were denatured in a solution containing 50% deionized formamide, 2 M formaldehyde, 20 mM morpholinopropane sulfonic acid (MOPS), pH 7.4, and 1 mM EDTA for five minutes at 65°C. Gels were 1.0-1.4% agarose in 1 M formaldehyde, 20 mM MOPS, pH 7.4, 1 mM EDTA, 0.5 g/ml ethidium bromide. The running buffer was the same as that for the gel but without formaldehyde. Gels were run At 50 volts overnight or 90 volts for six to eight hours. When mammalian and E. coli RNAs are run on the same gel, six ribosomal RNA bands are detectable and provide useful size markers. We use conversions of: 45S (13,700 nt), 32S (5800 nt), 28S (4700 nt), 18S (2000 nt) (137,163), 23S (2904 nt), and 16S (1541 nt) (70). Gels were photographed and the RNA was transferred to nitrocellulose filters (182).

Hybridization of Filters

cDNA plasmids pMA615 (albumin) (Dugaiczyk, unpublished) and pAF197 (AFP) (95) were provided by Dr. Achilles Dugaiczyk. Plasmids were nick translated according to the method of Rigby (147), using an enzyme reagent kit (Bethesda Research Laboratories). Nick translations using deoxycytidine 5'-[α-^{32}P]triphosphate (Amersham Corporation) routinely gave probes of specific activity greater than 5 x 10^8 cpm/μg. Unincorporated label was removed by passage through a column of Sephadex G50-fine. Probes were denatured immediately before use in hybridization, for five minutes in 0.3 M NaOH at 65°C, and then neutralized. Hybridization for thirty-six hours was a modification of that used by Thomas (182) with the exception that dextran sulfate was omitted and 0.1% SDS was added. Identical results were obtained using 50% formamide buffers at 42°C or aqueous buffers at 65°C. Filters were washed as in (182) except the last series of washes were done three times, with constant agitation, for forty-five minutes each at 52°C. Hybridized RNAs were detected by autoradiography at -90°C for one to five days using Kodak XAR film and a Chronex Lightning-Plus intensifying screen.

III. RESULTS

Analysis of Albumin RNA Intermediates of Processing

When citric acid is employed in the isolation of liver nuclei, the yield of nuclear RNA is 60-70 μg per gram of tissue. Using an NP-40 containing buffer, the yield averages about twice that obtained

with the citric acid buffer. This difference may result from extraction of some of the RNA by the citric acid or a contamination of the NP-40 nuclear RNA by cytoplasmic components. Both procedures enrich greatly for nuclear RNA since the recovery of total RNA is about 5-6 mg per gram of adult liver. When nuclear RNA is analyzed by hybridization of RNA blots with cDNA probes, results as shown in Figure II-2 are observed. The rRNA markers seen in Figure II-2, lanes A1 and B1, are characteristic of intact nuclear RNA. The autoradiograms show strongly hybridizing mature mRNA bands at about 2400 nt (range: 2113 to 2455) and multiple albumin specific bands that are mostly larger than the mature mRNA. This mRNA size is slightly larger than that reported using other techniques (32,161) and is consistent with the reported over-estimation of RNA sizes using RNA blotting methods (185). The size data presented for specific bands represent raw figures derived directly from standard curves obtained by plotting the log of the length of rRNA markers against the distance traveled in the gel. Because of factors such as broad RNA bands, variability in transfer and broad autoradiographic bands, significant size variation is observed even for bands of known identity (e.g., mature mRNAs). When the average size of a band is discussed a range of the observed sizes is also given. The experiment shown in Figure II-2A was done with RNA isolated from nuclei prepared in citric acid containing buffer. The blank areas seen in the high molecular weight range of this autoradiogram are artifacts of transfer. Figure II-2B shows a similar analysis of RNA from nuclei

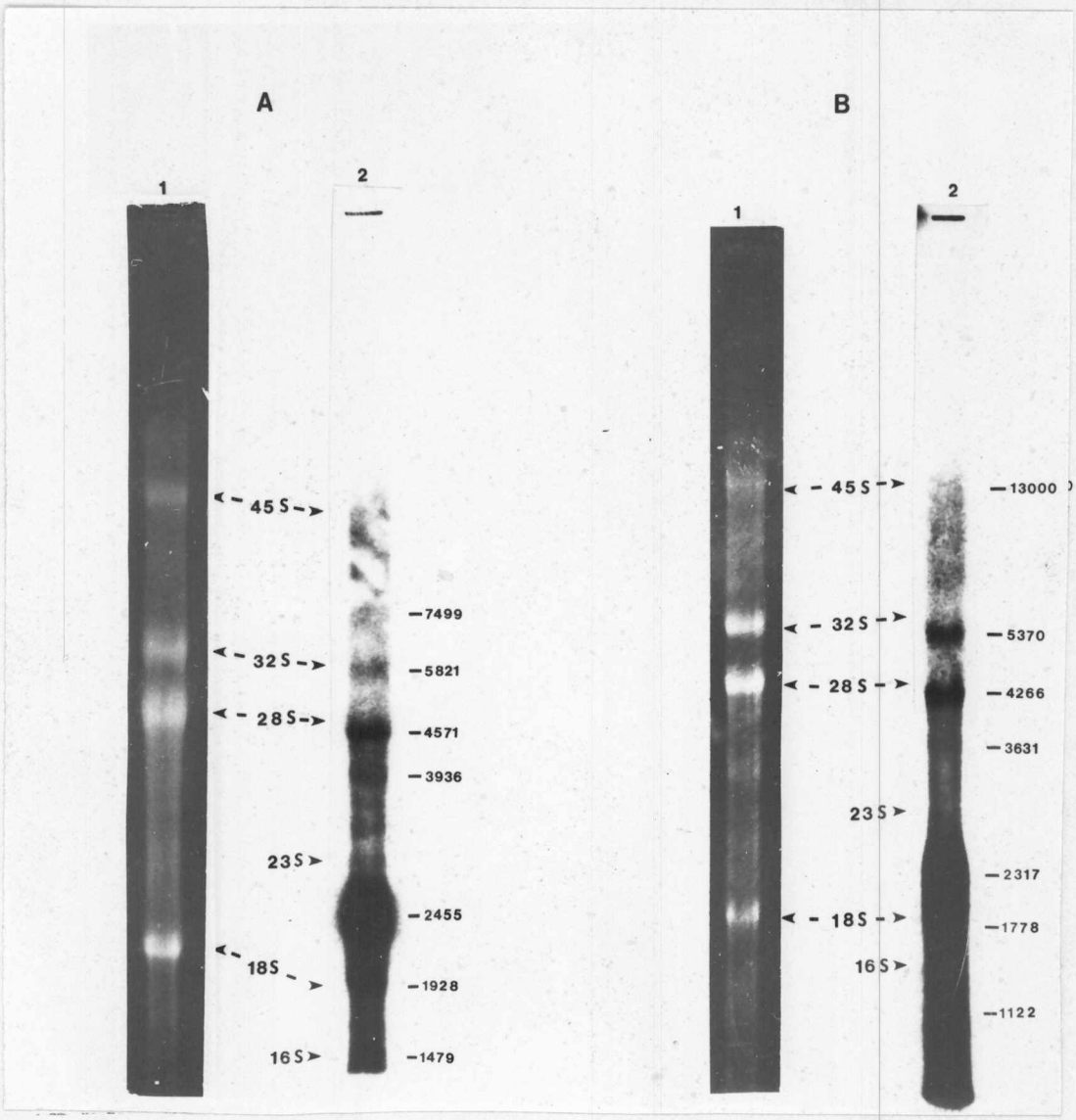


Figure II-2. RNA blot analysis of albumin nuclear RNA intermediates of processing.

In both A and B, lane 1 shows the electrophoretically resolved RNAs stained with ethidium bromide and photographed using an ultraviolet light source. Lane 2 shows the autoradiogram resulting from transfer of the RNAs in lane 1 and hybridization with pMA615 albumin cDNA probe. Sizes of significant RNA species are indicated in nucleotides and the rRNA standards are indicated. Mature mRNA bands are labelled 2455 and 2317 in A and B respectively. The experiment shown in A was done with 15 μg of RNA purified from nuclei prepared in citric acid. The experiment shown in B was done with 20 μg of RNA purified from nuclei prepared using NP-40.

prepared in NP-40 containing buffer but the resolution of individual precursor species is better in Figure II-2A. This slight loss of resolution is usually seen with the NP-40 technique and probably reflects its relative inefficiency at inhibition of ribonuclease activity (33). No one technique for preparing isolated nuclei was found to consistently yield undegraded nuclear RNA. The predicted 17 kb primary transcript for albumin (88) is undetectable using either technique. For albumin the largest detectable RNA precursor band is in the range of about 13,000 nt.

In repeated analyses, two RNA precursor bands of relatively strong intensity are detected but the sizes of these bands are variable. In the examples shown, the two intense precursor bands in the NP-40 experiment (bands labelled 5370 and 4266 in Figure II-2B, page 21) are not the same as the two strong bands seen in the citric acid autoradiogram (bands labelled 4571 and 3936 in Figure II-2A). Both of the autoradiograms clearly show a strong band that runs slightly smaller than the 28S rRNA precursor which is about 4700 nt in length. This strong band, with an average length of about 4400 nt (range: 4070 to 4571), is consistently seen as one of the two most intense precursor bands. The second band which gives a strong signal is either the 3936 nt band (range: 3673 to 3936) seen in Figure II-2A or the 5270 nt band (range: 5129 to 5821) seen in Figure II-2B. The observed differential intensity of these bands is an indication of the steady state levels of these RNA intermediates of processing. The variable signals of the high intensity bands may

therefore reflect a variability in the pathways of processing of these RNA intermediates. In contrast, consistency does exist because since in almost cases, two strongly intense bands are detected and in no case have strongly hybridizing bands been seen that are larger than the 32S ribosomal precursor.

In most of these experiments another set of two or more bands has been clearly observed that are smaller than the mRNA for albumin. The two most commonly observed small species have average sizes of 1800 nt (range: 1718 to 2065) and 1300 nt (range: 1122 to 1474) in length. These hybridization experiments have been done under stringent conditions and hybridization of the cDNA probe to nuclear RNA that does not contain albumin mRNA does not detect these species, therefore, they are considered to be albumin specific bands although neither a function for them nor a specific origin has been determined at this time (see Chapter IV). In similar experiments performed using probes for ovalbumin (153), immunoglobin light chain (136), and rat albumin (56), bands smaller than the mature message are also seen. The small sized bands were not discussed for either ovalbumin or immunoglobulin. In the rat albumin study, a small albumin specific RNA band was found in the normal nucleus and cytoplasm and it was suggested that it was not the result of RNA degradation. However, no alternative explanation of its origin was proposed (56).

Analysis of Nuclear RNA Fractionated on Oligo[dT]-Cellulose

Addition of a tract of poly[A] to the 3' end of mRNA is a modification known to occur very early after transcription (117). When nuclear RNA is separated into poly[A]-containing and poly[A]-lacking populations and analyzed as above, the results shown in Figure II-3 are obtained. Most of the detectable precursor molecules are in the poly[A]-containing sample. Although a small signal appears at a size corresponding to mature mRNA (2264 nt) in the poly[A]-depleted fraction, this is consistent with previously observed overlap of these populations (108) and with the observation that the 3' poly[A] sequences are gradually lost from RNAs as they age (162). The small 1800 nt albumin band is also detected in both the poly[A-] and poly[A+] fractions. Although poly[A] tails can be as long as 200 nucleotides and shorten with time, loss of poly[A] from the 2264 nt albumin mRNA cannot account for the observation of an 1800 nt band.

Comparison of RNA Processing for Albumin and AFP

It has been shown by many laboratories that AFP is evolutionarily related to albumin but that its expression is differentially regulated (168,183). Significant levels of AFP are produced only by fetal tissue or by liver tissue that has been damaged or undergone transformation. Because of the common ancestral origin and identical intron/exon organization of albumin and AFP (Figure II-1, page 15), attempts were made to detect and compare the RNA processing intermediates for AFP. The livers of newborn mice have been

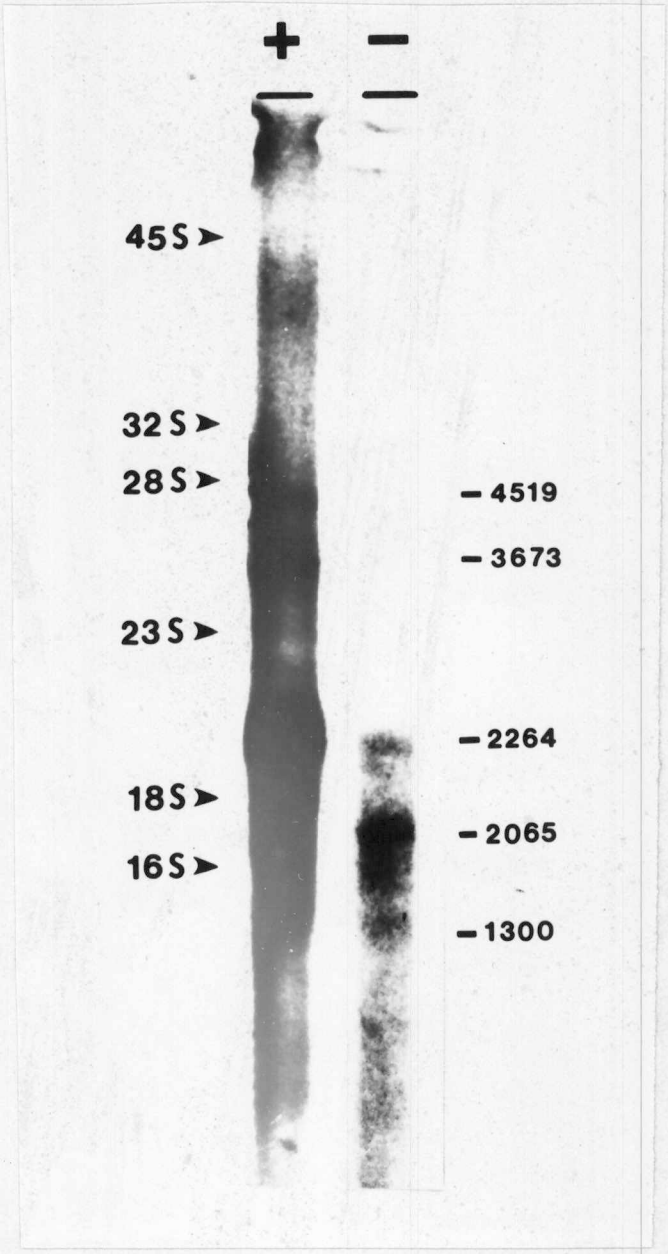


Figure II-3. Detection of albumin specific nuclear RNA molecules with and without poly[A].

Sizes of rRNA standards are shown to the left of the figure, while prominent RNA species are indicated in nucleotides on the right. Eighteen μg of RNA was analyzed in each lane and mature albumin mRNA is indicated at 2264 nucleotides.

shown to produce both albumin and AFP at high levels and were therefore expected to be a good source of RNA processing intermediates for both types of RNA. A second source of tissue synthesizing both of these proteins was the spontaneous minimal deviation hepatoma isolated at Oak Ridge (unpublished) and maintained via intramuscular transplantation in C3H mice. In order to ensure that the tumors were producing AFP, mouse serum from tumor bearing animals was tested using an agar double diffusion assay with antialphafetoprotein before isolating tumor nuclear RNA. Figure II-4 presents an experiment that compres the nuclear RNA precursors of adult liver and the two AFP producing tissues. The lanes shown in part A of this figure have been hybridized with albumin cDNA while part B was hybridized with AFP cDNA. The specificity of the AFP cDNA clone is demonstrated by the fact that this probe does not detect any of the adult RNA in lane B1. In lanes B2 and B3, analysis of RNA isolated from newborn liver nuclei and an AFP producing tumor, reveals a number of bands complementary to the AFP cDNA probe. The largest hybridization signal in lane B2 is beyond the accurate range of measurement using the rRNA standard curve but it is in the 18 kb range expected for a primary transcript of the AFP gene (88). The large overexposed band of about 2300 nt (range: 2113 to 2317) mRNA for AFP is similar in size to that for albumin. Both lanes of RNA precursors detected with the AFP probe (Figure II-4B, lanes 2 and 3) contain two relatively intense bands (labelled 4169 and 3467 in lane 2; and labelled 4365 and 3801 in lane 3). The

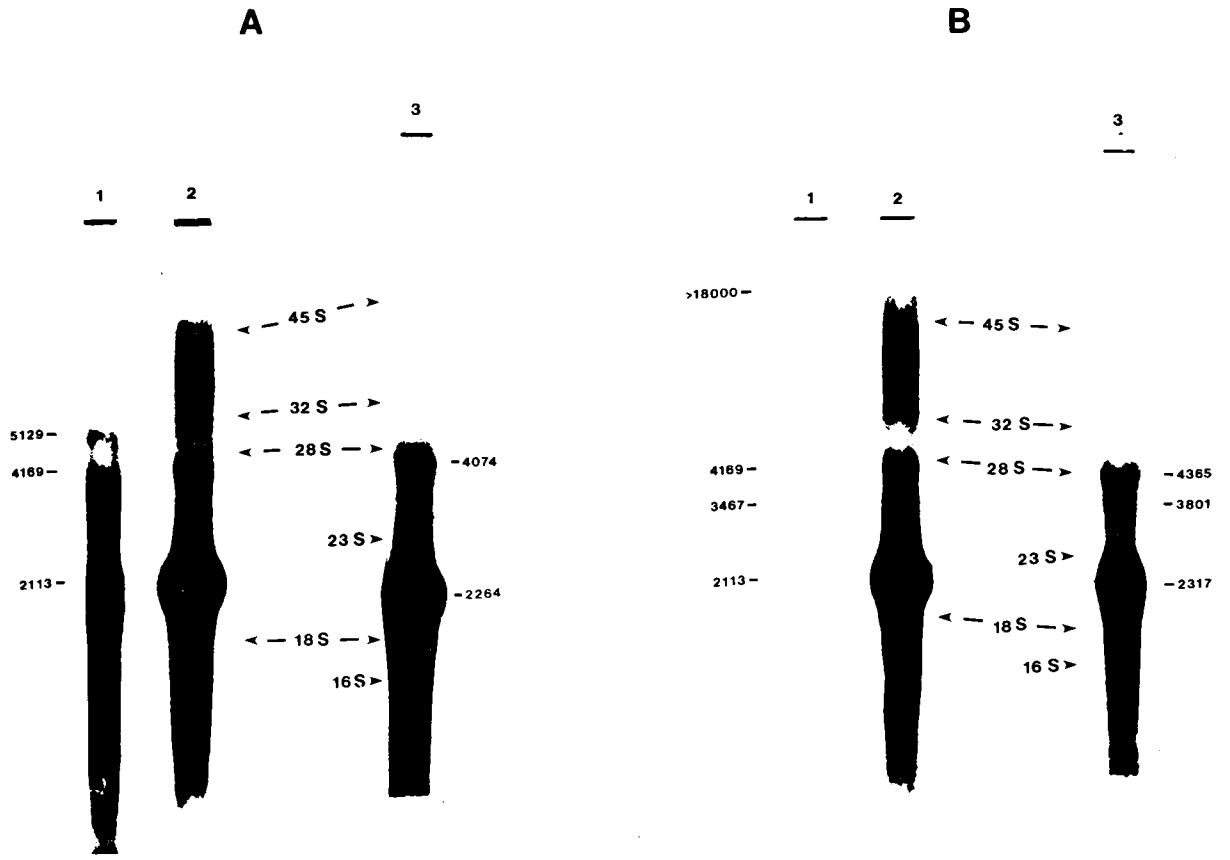


Figure II-4. Comparison of nuclear RNA precursors for albumin and AFP.

Lanes 1, 2, and 3 represent blot experiments done using RNA isolated from adult liver, newborn liver, and hepatoma tissue respectively. The blots shown in A were hybridized with albumin cDNA and those in B were hybridized with AFP cDNA. Lanes 1 and 2 are samples run in the same gel and, therefore, the same sizes of indicated RNA bands and rRNA standards apply to both. Lanes labelled 3 in both A and B were run in separate gels and so have a slightly different scale of rRNA standard locations.

observation of two strongly hybridizing bands of AFP RNA precursors is similar to that of two strongly hybridizing bands in the albumin RNA precursors shown in Figure II-4A, lane 1 (bands labelled 5129 and 4169). In both, the albumin precursors seen in adult nuclear RNAs and the AFP RNA precursors, at least two strongly hybridizing bands are always detected. Albumin synthesis also occurs in both of the tissues that synthesize AFP. When nuclear RNAs from the AFP producing tissues are analyzed for albumin precursors, instead of detecting two bands at 5129 nt and 4169 nt as in the adult RNA sample (Figure II-4A, lane 1), only a single strongly hybridizing band is evident (labelled 4169 in lane A2 or 4074 in lane A3). Thus, there exists an alteration in the bands of albumin nuclear RNA precursors, which is characteristic of tissues that are also producing AFP.

Any study that attempts to describe the sizes of specific RNAs must consider the possibility of artifacts that have been introduced by ubiquitous RNase activity. The alteration in the albumin RNA precursor bands detected in AFP producing tissues is believed not to be a result of degradation: first, because the 455 RNA precursor bands are detectable with ethidium bromide staining in all of these gels. Second, in the case of newborn RNAs (Figure II-4B, lane 2), analysis of the AFP RNA intermediates consistently shows a large molecular weight RNA species and two intense intermediates typical of results seen with intact RNA.

IV. DISCUSSION

Because the first and last exons of AFP and albumin contain the cap site and poly[A] addition signal respectively (55), it is conceivable that sequences outside of these landmarks might not be included in the products of transcription of these genes. In most cases that have been studied, the cap site does indeed begin the RNA sequence of genes transcribed by RNA polymerase II (77). The 5' ends of the five early transcription units of adenovirus all begin at their respective cap sites, with a variation of 2-6 bases, in some cases (200). Transcription of the β-globin gene in Friend erythroleukemia cells has also been shown to begin at or very close to the cap site in at least 80% of the RNA molecules examined (76). It seems likely from this type of evidence that transcription of AFP and albumin would also begin at the first exon.

The 3'-end of eukaryotic messengers, however, is not as easily defined. Even though a common polyadenylation signal has been identified at the 3'-end of many genes (143), transcription in some cases continues thousands of nucleotides beyond this point (76). For the adenovirus late transcript no less than five poly[A] signals are passed before the RNA polymerase reaches the end of the genome (58,118). In the globin system, several reports have appeared of globin-specific RNAs greater than 5000 nt in length in addition to a polyadenylated 1800 nt precursor that contains all of the coding and noncoding regions (12,154,176). Factors which might regulate

the termination of transcription in eukaryotic genes are largely unknown. For histone genes whose RNA transcripts are not polyadenylated, a conserved hairpin loop structure (173) and a ribonuclear protein particle (59) seem to be involved in normal termination. For eukaryotic genes that code for other proteins, it may be the composition of the regional chromatin that regulates these events (77).

For albumin and AFP synthesis of a primary RNA transcript of at least 17 and 18 kb, and possibly longer, would be expected in order to contain all of the exon and intron regions (88). In the experiments described here for albumin nuclear RNA, precursors larger than about 13,000 nucleotides have not been detected and those larger than about 6,000 nucleotides are poorly resolved and exist in low abundance (see Figure II-2, page 21). Early studies of albumin RNA precursors, in which a less sensitive sucrose gradient separation of nuclear RNAs was used, revealed only molecules of 28S and smaller (177). These data imply that processing steps affecting the large albumin RNA precursors occur relatively rapidly when compared to the more easily detectable smaller RNA intermediate species. A report of detected albumin RNA precursors in analbuminemic rats is difficult to compare with the results shown here because a splicing defect is described and the RNA precursors detectable in normal liver are not presented (56). The detectable set of albumin RNA precursors is shown in Figure II-3, page 25, to be polyadenylated as would be expected if a poly[A] tract is added early

after transcription, before RNA processing has begun. The early addition of poly[A] has been demonstrated for RNA precursors of immunoglobulins (151), adenovirus (117), and growth hormone (69). The universality and early addition of poly[A] to RNA precursors transcribed from split genes would seem to argue for some important function for this sequence even though RNA processing occurs without this sequence and therefore is not required for this process (199).

For AFP, a low abundance species of RNA has been detected in nuclear RNA from newborn mouse liver (see Figure II-4B, lane 2) that may be large enough to contain all of the regions of the gene. As expected, no AFP sequences were detected in nuclear RNA from adult liver which formally discounts the unlikely possibility that re-expression of AFP during liver damage is effected through processing of previously synthesized and stored mRNA precursors. In other systems, such as ovalbumin (153), ovomucoid (119), immunoglobulins (138), growth hormone (105), and prolactin (7887), polyadenylated RNA precursors large enough to be a primary transcript have been detected. The largest of these precursors is 14,000 nt for prolactin (78). The precursor molecules detected here are, to our knowledge, the largest that have been reported using this technique. Although formal proof of the precursor relationship of these RNA intermediates has not been presented, the large body of available data supporting precursor functions for these kinds of nuclear RNAs make other interpretations highly unlikely (28).

The species of RNA precursors seen in these experiments represent those existing in a detectable steady state concentration. Two of these RNA precursor bands are detectable at consistently greater concentrations than the other precursor bands for both AFP and for albumin in adult nuclear RNAs. These bands are interpreted to result from rate limiting steps in the processing pathway. The detection of this similar result might be expected for two such closely related genes, however, the fact that the sizes of these bands show such concordance may be fortuitous since the sizes and sequences of the analogous introns have significantly diverged (88). Variations in sizes of detected RNA intermediates for these genes is partially a result of inherent irreproducibility of the RNA gel electrophoresis and transfer technique but may also result from variability in the sequence of intron removal. Many of the introns in the two genes are very similar in size (88) and therefore different orders of removal of these regions from nuclear RNAs could cause subtle changes in sizes of detected intermediates. Experiments designed to further analyze the order of removal of specific introns from RNA precursors are presented in Chapter IV.

It has long been recognized that tissues that have undergone carcinogenic transformation often resemble the fetal phenotype in their protein expression patterns (166) and many of these so-called oncofetal antigens have been studied. AFP is a frequently cited example of such a protein and its synthesis represents one example of a fetal function seen in hepatomas (166). In Figure II-4, page 27, the newborn mouse liver and hepatoma tissues that are shown to

contain AFP RNA precursors also contain albumin RNA sequences. The single intense band of albumin RNA precursor seen here however (Figure II-4, lanes 2 and 3, page 27) has only been seen when analyzing nuclear RNA from tissues that are also producing AFP. This observation might be interpreted as a change in the rate limiting steps in the processing of albumin nuclear RNA. When an albumin RNA transcript is being synthesized in a tissue that also produces AFP, its intron splicing signals have not changed. If the proposed limited set of splicing enzymes has also not changed (36), then perhaps different sets of hnRNP proteins are applied to nascent albumin transcripts in these tissues. The actual significance of this observed phenotypic variation is not immediately obvious.

Albumin specific RNA bands that are smaller than mature mRNA sequences have been seen in these studies for both albumin and AFP and, at least for albumin, they are present in both the poly[A+] and poly[A-] fractions. The data presented in this chapter does not allow an interpretation of the origins or function of these species. Further discussion of them can be found in Chapter IV.

CHAPTER III

RESTRICTION ENZYME MAPPING AND SUBCLONING OF ALBUMIN INTRON SPECIFIC REGIONS

I. INTRODUCTION

Knowledge of the structure of the albumin and AFP gene locus in the mouse has been limited to the large scale maps published by Tilghman's group (63,79,87). In addition, sequence data are available for the AFP gene 5' and 3' regions (165), the exon/intron boundaries (55), and mRNA (cDNA) (62). Structural information for the human (53,71,111) and rat genes have also been published (82, 160,161). In order to conduct further analysis of regulatory events involving these genes, a more detailed characterization was needed. This is especially true for those studies where additional knowledge of the intron/exon organization of these genes might reveal features exploitable in the analysis of RNA processing steps. In addition, it was anticipated that, as in other systems, structural features might be revealed that are involved in the unique regulation of this system. Among the known functional regions of the eukaryotic gene are those such as the TATA sequence that specifies transcription start sites (promoter region) and poly[A] addition sites (143). In the promoter region, sequences that appear to be required for proper regulation by steroid hormones have been found (113). Recent studies indicate that these are similar to the enhancer sequences

observed in viruses. For example in MMTV (mouse mammary tumor virus), which is regulated by glucocorticoids, the glucocorticoid-receptor protein complex has been shown to bind to these enhancer sequences (131).

A number of families of repeated DNA sequences have also been observed around eukaryotic genes as well as in intergenic regions. In the globin system, several different types of repeated DNA sequences have been located interspersed between related members of this multigene family (67,171). Repeated sequences have also been detected in the 5' flanking region and in one intron of the conalbumin gene (41) and in rat growth hormone (127). Repeated DNA regions have been detected in association with many other genes (49) but a function for them remains speculative (125). In the AFP and albumin genes, the locations of related repeated sequences have been reported in the 5' region of AFP, in intron 3 of AFP, and intron 12 of albumin (88). The original publication placed the intron 12 repeat in intron 3 of albumin but a correction reversing the map direction relocalized the repeat to intron 12 (79). In addition, an Alu-like repeat has been sequenced from intron 1 of AFP (165). This type of repeated sequence is very common in the mouse genome (148) and is related to the highly reiterated Alu sequence in humans (90,164). More surprising than its unexpected location, is the transcription of this sequence into a 340 nt RNA from the opposite DNA strand (196). As with other Alu-like repeats, poly(dA) tracts, Polymerase III start signals, and direct repeat flanking

regions are seen (196). Characterization of the albumin gene was begun by restriction enzyme mapping studies of subcloned albumin Eco RI gene fragments (see Figure II-1, page 15). These experiments were designed to confirm and extend the available published data. cDNA hybridization experiments were then performed to identify which of the known restriction fragments were uniquely of intron origin. Four fragments that were shown to be free of exon regions were further subcloned into the plasmid vector pBR322 and amplified for use as probes. Analyzing for the uniqueness of these sequences revealed that one of the four probes contained a repeated sequence. Results of mapping and characterization of these albumin gene regions are presented in this chapter.

II. EXPERIMENTAL PROCEDURES

Mapping Albumin Genomic Clones

Charon 4A (25) lambda clones of albumin genomic DNA sequences were provided by Dr. James Bonner. These clones were created from sheared mouse plasmacytoma genomic DNA and by ligation with linkers prior to their insertion into Charon 4A. The two clones λMSA1 and λMSA3 were mapped and found to contain a number of EcoRI fragments, four of which together spanned the entire albumin gene (unpublished). These four fragments were subcloned into the EcoRI site of pBR325 and are named for the lambda clones from which they are derived. The 5' to 3' order in which these Eco RI subclones span the albumin gene is 3-1, 3-2, 1-3, and 1-5 (see Figure II-1, page 15).

Digestion with restriction enzymes was according to the manufacturer's (Bethesda Research Laboratories) specifications. DNA gels were composed of 1.2% agarose in a buffer containing 40 mM Tris-HCl, pH 7.7, 5 mM sodium acetate, and 1 mM EDTA. Size standards were generated from lambda phage digested with Hind III (159) and from pBR322 (179) digested with either Alu I, Ava II, or Eco RI and Pst I. Fragments separated by gel electrophoresis were transferred as described by Southern (175) and Wahl et al. (188). Hybridizations to nitrocellulose filters were as described by Wahl et al. (123) but without glycine and dextran sulfate. Washing of hybridized filters was as described for RNA filters in Chapter II. The transfer and hybridization experiments shown in the figure on page 49 were performed using Gene Screen membrane (New England Nuclear) and by following the procedures described in the protcol provided by the manufacturer.

Subcloning of Intron Sequences

Fragments of DNA were isolated from agarose gels by electrophoretically transferring the band of interest onto a piece of diethylaminoethyl (DEAE) paper. DNA was eluted from this paper by two 50-minute incubations at 50°C in a buffer of 1 M NaCl, 10 mM Tris-HCl, pH 7.4, and 1 mM EDTA. The elution solution was separated from the paper by centrifugation through a small hole punched in the bottom of the incubation tube (194).

In all cases of intron subcloning, the fragment of interest was enriched by electrophoresis and isolated from the gels as

described above. This procedure served to simplify the selection of resulting recombinants needed for these studies. The intron 6 insert is an 825 bp fragment that results from digestion of the 3-2 albumin subclone with Hind III (see figure on page 45). This fragment was ligated to pBR322 that had been linearized at the Hind III site. Ligation reactions were done at 15°C for 16-24 hours in 67 mM Tris-HCl, pH 7.4, 67 mM MgCl, 10 mM dithiothreitol, 0.4 mM ATP, and 25 units per milliliter of T4 DNA ligase (Amersham Corporation).

The intron 7 insert was prepared by digestion of the 3-2 subclone with Kpn I and Sst I and the resulting 2900 bp fragment was isolated from an agarose gel as described above (see page 45). This fragment was subsequently digested with Msp I and the 450 bp fragment produced was also isolated by agarose gel electrophoresis. This 450 bp fragment was ligated into pBR322 that had been opened with Cla I. Cla I and Msp I ends will ligate but will not recut, therefore recombinant inserts were screened by digestion with both EcoRI and BamHI. This digestion releases a 373 bp fragment if there is no insert, and an 823 bp fragment if there is an insert.

The intron 10 insert is the 288 bp fragment that results from digestion of the 3-2 subclone with Hind III and Eco RI (see page 45). It was ligated into pBR322 that had been digested with Hind III and EcoRI and then purified by gel electrophoresis in order to lower the background of intact pBR322 on antibiotic selection plates.

The intron 12 insert is the 600 bp fragment that results from digestion of the 1-3 subclone with Pvu II and Bam HI (see the figure on page 47). This insert was ligated with the larger of the two fragments that result from the digestion of pBR322 with Pvu II and Bam HI.

The RR1 strain of E. coli was transformed (46) with the products of the above ligations and plated onto selective medium containing 50 μg/ml ampicillin. Tetracycline sensitive colonies were selected and then further screened by restriction analysis of alkaline lysates (24) (see summary, Table III-1). Plasmid DNA was prepared as described (40,120) using CsCl/ethidium bromide gradients.

Nick Translation of Genomic DNA and Plasmid Clones

Total Balb/c liver genomic DNA was digested with Eco RI and Hind III and then phenol extracted and precipitated with ethanol. This material and all plasmid clones were nick translated as described in Chapter II.

Genomic Blots

Blots of genomic DNA were prepared and hybridized as described above for the mapping gels. About 7 μg of EcoRI digested genomic DNA was applied to each lane.

III. RESULTS

Restriction Enzyme Mapping of Albumin Subclones

Four Eco RI genomic subclones inserted into pBR325 were chosen for analysis. These subcloned fragments are found in albumin

TABLE III-1

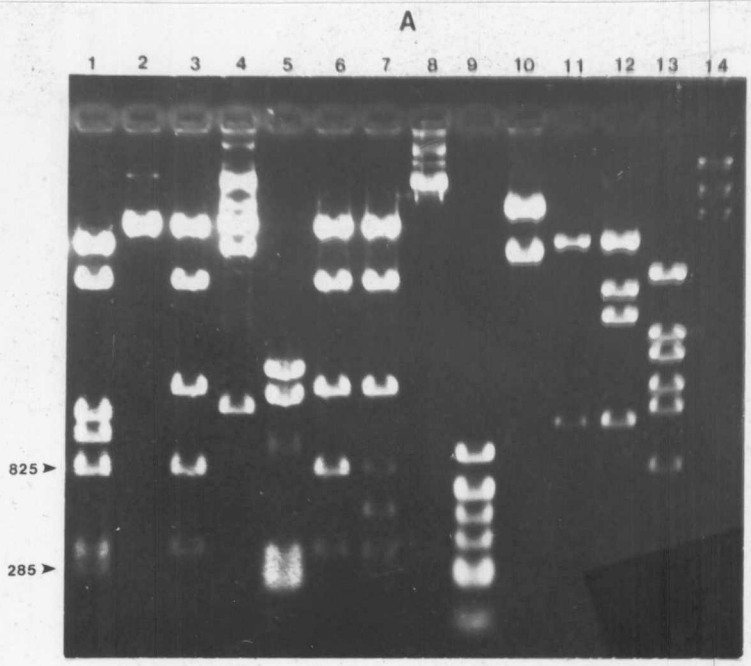
SUMMARY DATA FOR INTRON SUBCLONES

DNA Fragment ID	Insert Boundary Restriction Sites 5'-3	Insert Size in Base Pairs	Selection Phenotype	Occurrence in Mouse Genome
I-6	Hind III-Hind III	800	Amp^r, Tet^s	Unique
I-7	Msp I-MspI	450	Amp^r, Tet^s	Unique
I-10	Hind III-Eco RI	288	Amp^r, Tet^s	Unique
I-12	Pvu I-Bam HI	595	Amp^r, Tet^s	Repeated

gene in the 5' to 3' order, 3-1, 3-2, 1-3, and 1-5 (see Figure II-1, page 15).

The 3-2 plasmid containing parts of introns 5-10 and including exons f-j was digested with various combinations of restriction enzymes based on published data (88) and other enzymes that were of interest. The location of Msp I sites was important because these sites are susceptible to methylation and this modification has been shown to be involved in the regulation of expression of several genes (144). Figure III-1A is a photograph of an ethidium bromide stained gel that demonstrates the results of this type of analysis. Many of the bands in these digestions represent fragments of pBR325, which has been sequenced, and so act as internal size standards. Lengths in base pairs were calculated for the products of each digestion and the data (see Table III-2 and Table III-3) from overlapping combinations of enzymes were used to construct a map of the plasmid and its insert. A map deduced from the plasmid digests shown in Figure III-1 and others is shown in Figure III-2. This type of analysis was performed for the other three Eco RI subclones also and the resulting maps are presented in Figures III-3 and III-4 (see also summary data in Table III-3). The exons on these maps cannot be precisely placed without sequence data and so represent the best placement from the available data.

A cDNA clone is a DNA copy of the sequences present in a mature mRNA molecule and as such represents only exon regions. The cDNA insert from the pMA615 plasmid, containing exons c through 1,



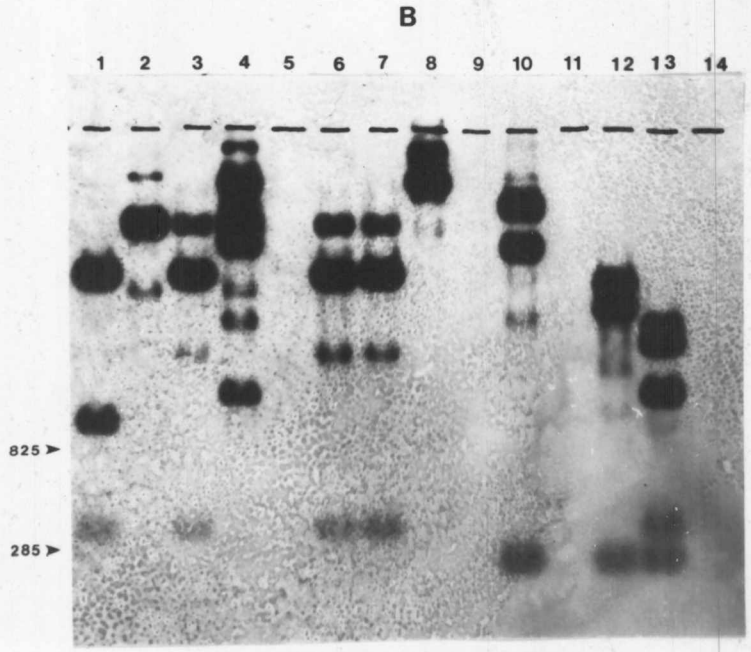


Figure III-1. Restriction digestion and Southern analysis of albumin Eco R1 subclone 3-2.

Part A is the ethidium bromide stained gel of digestion products. Part B is the autoradiogram that results from transfer to filter paper of the gel shown in A and hybridization with the exon specific (cDNA) probe pMA615. Lane numbers are the same in parts A and B and represent individual restriction enzyme digests of DNA which are identified in Table III-2. Two example fragments labelled 825 base pairs (bp) and 285 bp are identified in lane 1 which do not contain exon regions.

TABLE III-2

RESTRICTION ENZYME[a] FRAGMENTS OBTAINED BY DIGESTION OF 3-2 SUBCLONE AND SHOWN IN FIGURE III-1

Band[b] Number	H+E (1)	E (2)	H (3)	E+K (4)	(pBR322)[c] Av (5)	H+S (6)	H+K (7)	K (8)	(pBR325)[c] Al (9)	P (10)	(pBR325)[c] E+P (11)	E+P (12)	H+P (13)	(λ phage)[d] H (14)
1	4749[e]	5956	5130	5956	1743	5130	5130	12589	910	6600	4829[e]	4829[e]	3583[e]	23100
2	3388		3388	4266	1433	3388	1550		659	4074	1166[e]	3020	2200	9416
3	1246[e]		1550	1350	345	1550	813		521	260		2570	1950	6682
4	1059		813		279[f]	813	550		403			1166[e]	1550	4361
5	813		359			359	359		281[f]				1320	2322
6	359						295		136[f]				900	2027
7	280												359	564
8													260	

[a]Restriction Enzyme Abbreviations: Al, Alu I; Av, Ava II; E Eco RI; H, Hind III; K, Kpn I; P, Pst; S, Sac I.

[b]Minor or partial digestion bands are not included in the data list; lengths of the DNA fragments shown in Figure III-1, page 42, are given in base pairs.

[c]pBR322 (179) plasmid and pBR325 (142) published values.

[d]λ phage DNA (Bethesda Research Laboratories) published values (159).

[e]Indicates bands derived from pBR325 vector (see note c).

[f]Indicates the largest band of an unresolvable group.

TABLE III-3

SUMMARY[a] RESTRICTED FRAGMENT DATA FOR THE ALBUMIN GENE

Eco RI Subclone 3-1[b]		Eco RI Subclone 3-2		Eco RI Subclone 1-3		Eco RI Subclone 1-5	
Restriction[c] Fragment	Fragment Length and Total[d]	Restriction Fragment	Fragment Length and Total	Restriction Fragment	Fragment Length and Total	Restriction Fragment	Fragment Length and Total
E to A	1050	E to H	1100	E to B	1480	E to A	400
A to H	100	H to K	270	B to B	875	A to Pv	530
H to A	855	K to H	555	B to S	280	Pv to H	280
A to H	1600	H to M	495	S to Pv	320	H to E	305
H to Sm	1800	M to M	450	Pv to P	550		1515
Sm to K	885	M to P	300	P to E	580		
K to E	1260	P to P	200		4085		
	7550	P to S	890				
		S to H	1040				
		H to H	370				
		H to E	285				
			5955				

[a]Average size data from multiple experiments rounded off to the nearest 5 bp.

[b]The four columns present the restriction fragment sizes that are presented in the maps in Figures III-2, III-3, and III-4, pages 45, 46, and 47 respectively.

[c]Restriction Enzyme Abbreviations: A, Ava I; B, Bam HI; M, Msp I; Pv, Pvu II; Sm, Sma I, and as in Table III-2. Fragments presented in 5'-3' orientation.

[d]Lengths given in base pairs. Totals represent the size of the Eco RI fragment subcloned into pBR325.

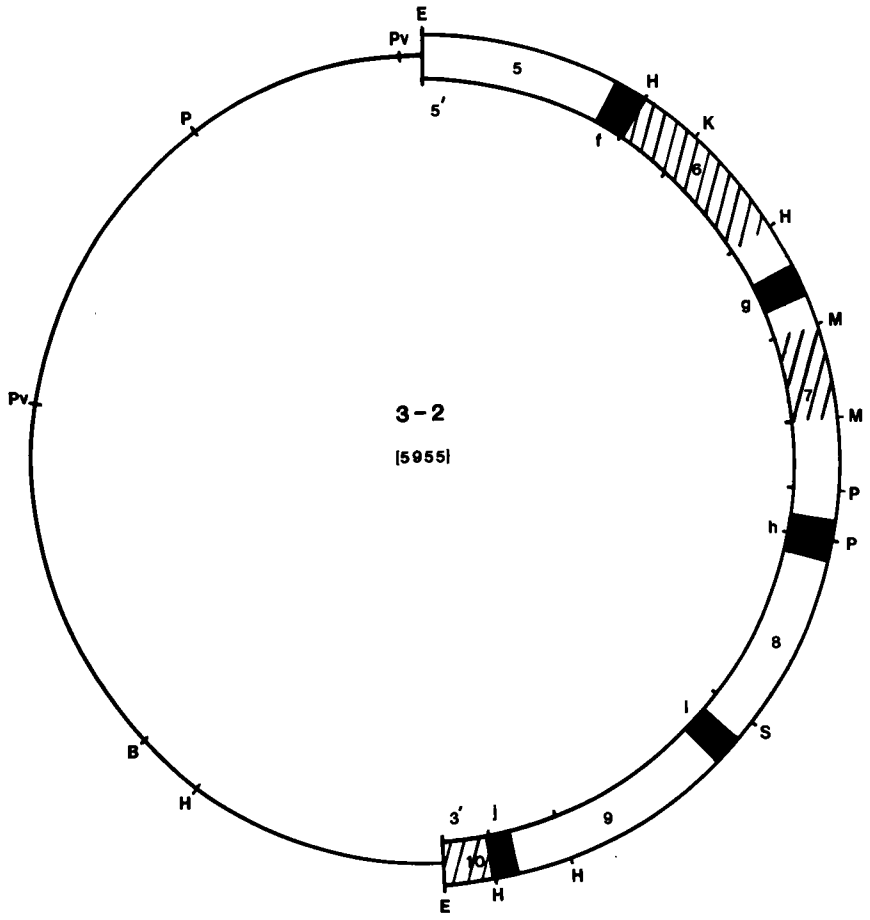


Figure III-2. Restriction enzyme map of albumin Eco RI subclone 3-2.

The single line represents pBR325 sequences. Double lines are albumin gene regions with exons shown in black boxes and introns shown in open boxes with their numbers. Exons are labelled in lower case letters inside the circle. Restriction enzyme sites are indicated at the outside of the circle and sizes and key to abbreviations can be found in Tables III-2 and III-3. Cross-hatched areas indicate intron regions that have been subcloned into pBR322. Size of the albumin gene region insert is shown in the center in parenthesis below the subclone numbers.

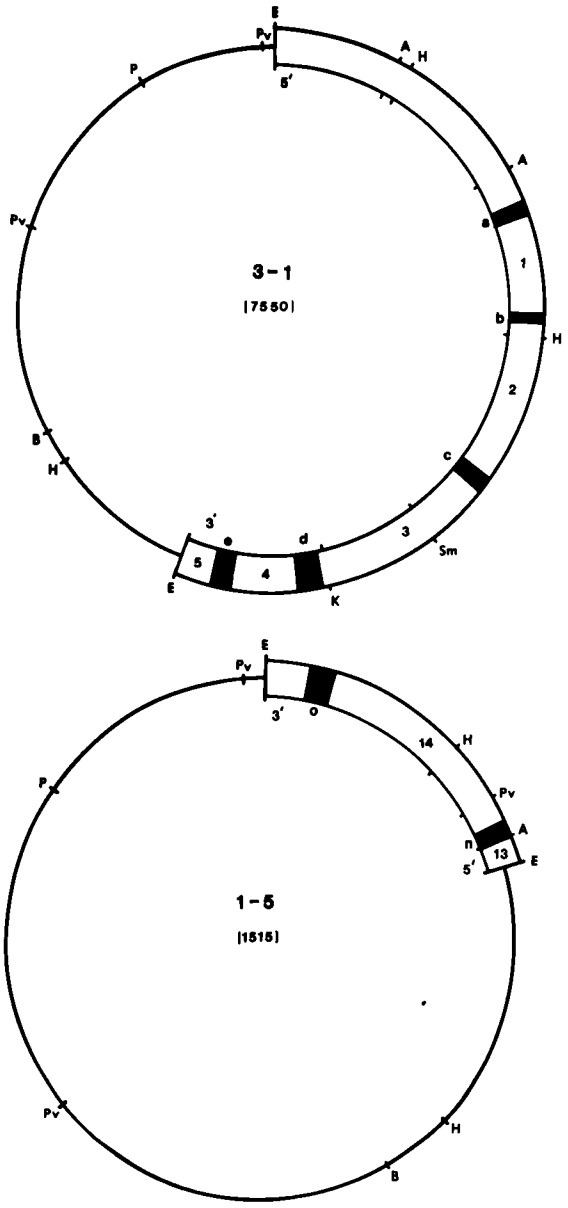


Figure III-3. Restriction enzyme map of albumin Eco RI subclones 3-1 and 1-5.

Open box areas without numbers represent albumin gene regions outside of the area included between exons a through o. See legend for Figure III-2.

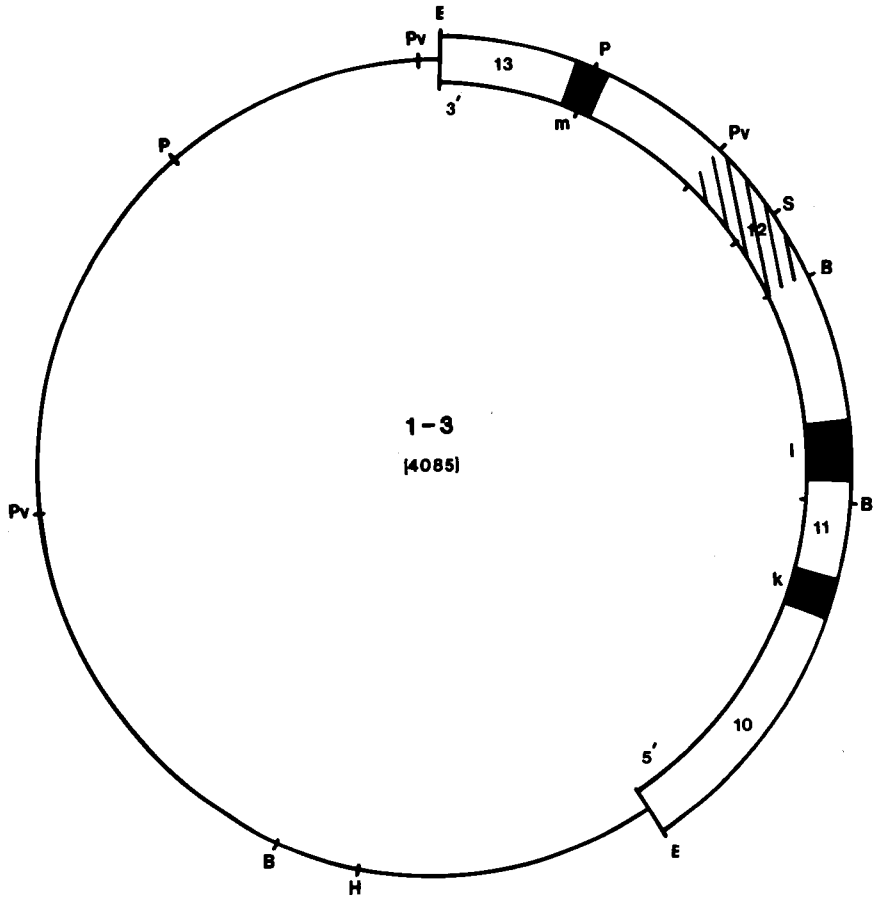


Figure III-4. Restriction enzyme map of albumin Eco RI subclone 1-3. See legend for Figure III-2.

was isolated on preparative agarose gels. The recovered insert was hybridized to DNA blots made by transferring the DNA from the mapping gels discussed above to nitrocellulose filters. When the gel shown in Figure III-1A, page 42, was analyzed in this manner, the autoradiogram shown in Figure III-1B was obtained. By comparing the pattern of bands in the photograph with those that give a positive signal in the autoradiogram, one can determine that there are several non-pBR322 fragments that do not contain exon regions. Two such fragments are identified in lane 1 of Figure III-1. This type of data, when combined with the map locations of the exon regions, allowed the assignment of intron specific DNA to three of the fragments shown on the map in Figure III-2, page 45. These fragments are derived from introns 6, 7, and 10. A fragment from intron 12, characterized in similar (see Figure III-5A and 5B) fashion is shown in Figure III-3.

Subcloning of Intron Fragments

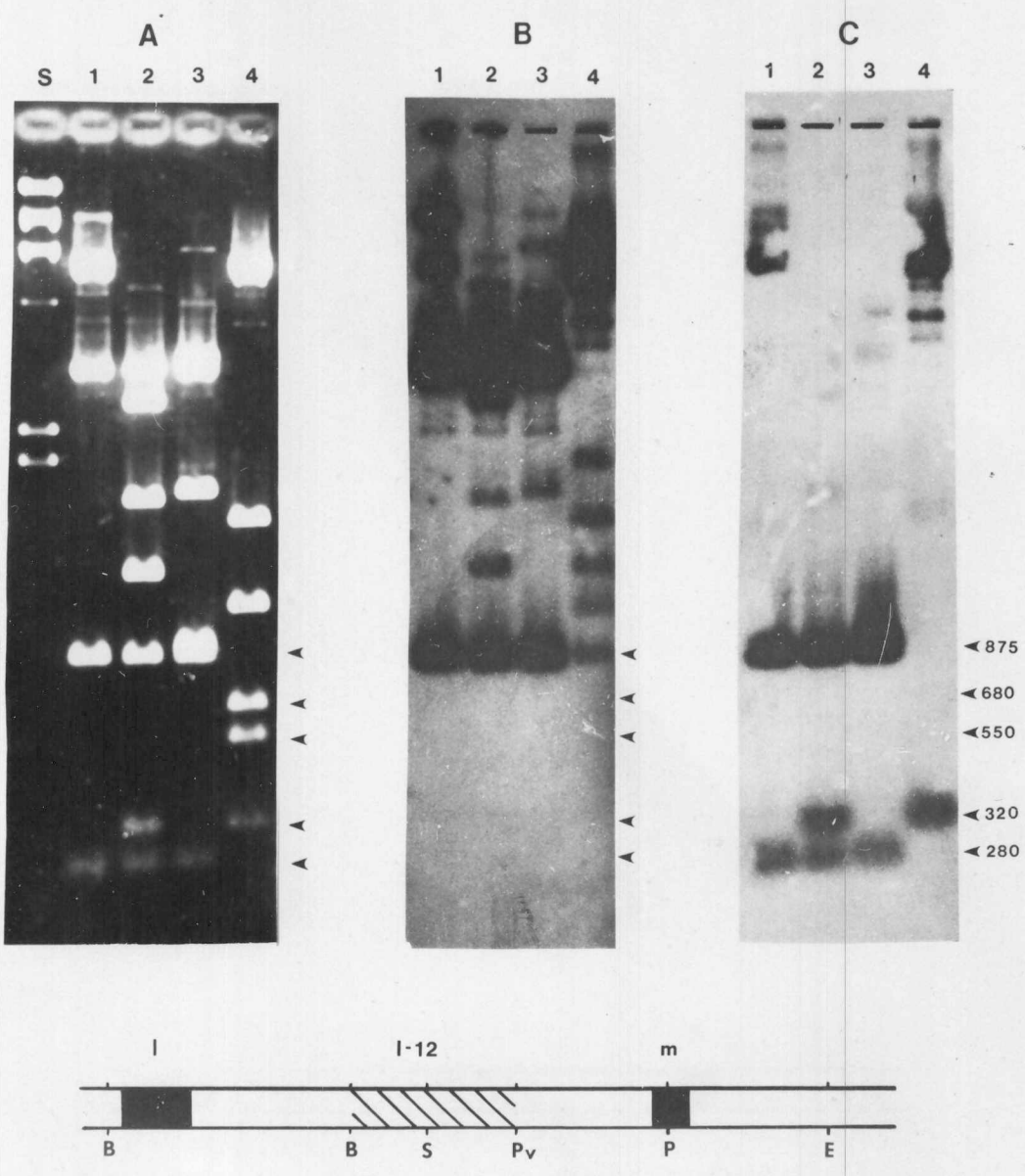
The four fragments from albumin introns 6, 7, 10, and 12 were isolated from a preparative agarose gel. Plasmid vector preparations were digested with the appropriate restriction enzymes to provide ends compatible for ligation with the intron fragments. Ligations, transformation, screening, and amplification of the intron subclones are described in the experimental procedures.

Repeated Sequence Analysis

Because repeated sequences had been reported in introns from other systems (41) and in other introns of this system (88), the

Figure III-5. Restriction enzyme digestion and Southern analysis of albumin Eco RI subclone 1-3.

The ethidium bromide stained pattern from several restriction enzyme digestions of Eco RI subclone 1-3 are shown in A. Duplicate gels as shown in A were transferred to filter paper and hybridized with pMA615 cDNA clone (autoradiogram shown in B) or a labelled total genomic DNA probe (autoradiogram shown in C). Lane S is the size standard pattern of λ phage DNA digested with Hind III (sizes given in Table III-2). Restriction enzymes used for digestions were: Lane 1, Sac I and Bam HI; lane 2, Sac I, Bam HI, and Pvu II; lane 3, Sac I, Bam HI, and Pst I; lane 4, Sac I, Pst I and Pvu II. The indicated fragments at 320 bp and 280 bp are shown in B not to contain exon regions and in C to hybridize to the genomic probe indicating the presence of repeated sequences. The 875 bp fragment is shown to contain exon sequences (B) and repeated sequences (C). The gene region shown at the bottom of the figure represents the relevant area around intron 12 also shown in Figure III-4. Two fragments labelled 680 bp and 550 bp do not contain repeated sequences and also show that the pMA615 cDNA probe does not include exon m.


A
S 1 2 3 4
B
1 2 3 4
C
1 2 3 4
875
680
550
320
280
I
I-12
m
B
B
S
Pv
P
E

intron subclones were used in hybridization experiments against total genomic DNA that had been digested with a restriction enzyme and blotted onto a nitrocellulose filter. If an intron fragment contained a sequence that was found only in the albumin region then it should hybridize to only a single fragment of genomic DNA that is predictable from the maps. If other sequences in the genome were homologous to the probe, then hybridization would occur everywhere those sequences occurred in the digestion pattern. The results of hybridizing these probes with EcoRI digested RNA is shown in Figure III-6. For comparison, a lane hybridized to the cDNA probe, pMA615, is shown in lane 1. Two predicted fragments at 6.0 kb (see Figure III-2, page 45) and 4.0 kb (see Figure III-4, page 47) are detected by the cDNA probe. The third cDNA detected fragment, at 10 kb, is larger than the 7.5 kb fragment in the 3-1 Eco RI subclone (see Figure III-3, page 46) because the 5' Eco RI site is an artifact of the original cloning. With an EcoRI enzyme digestion, only the middle fragment would be expected to be detected by the probes for introns 6, 7, and 10. They all detected the predicted 6.0 kb fragment. The result is consistent with the intron sequence being unique to the albumin gene. A weakly hybridizing band at about 5 kb is an artifact produced by nick translated pBR322 sequences since if only isolated insert fragments are nick translated, this signal is not seen (data not shown). Intron 12 would be expected to hybridize only to the 4 kb Eco RI fragment if it were also unique, but instead it hybridizes to a continuum of fragments. This pattern is

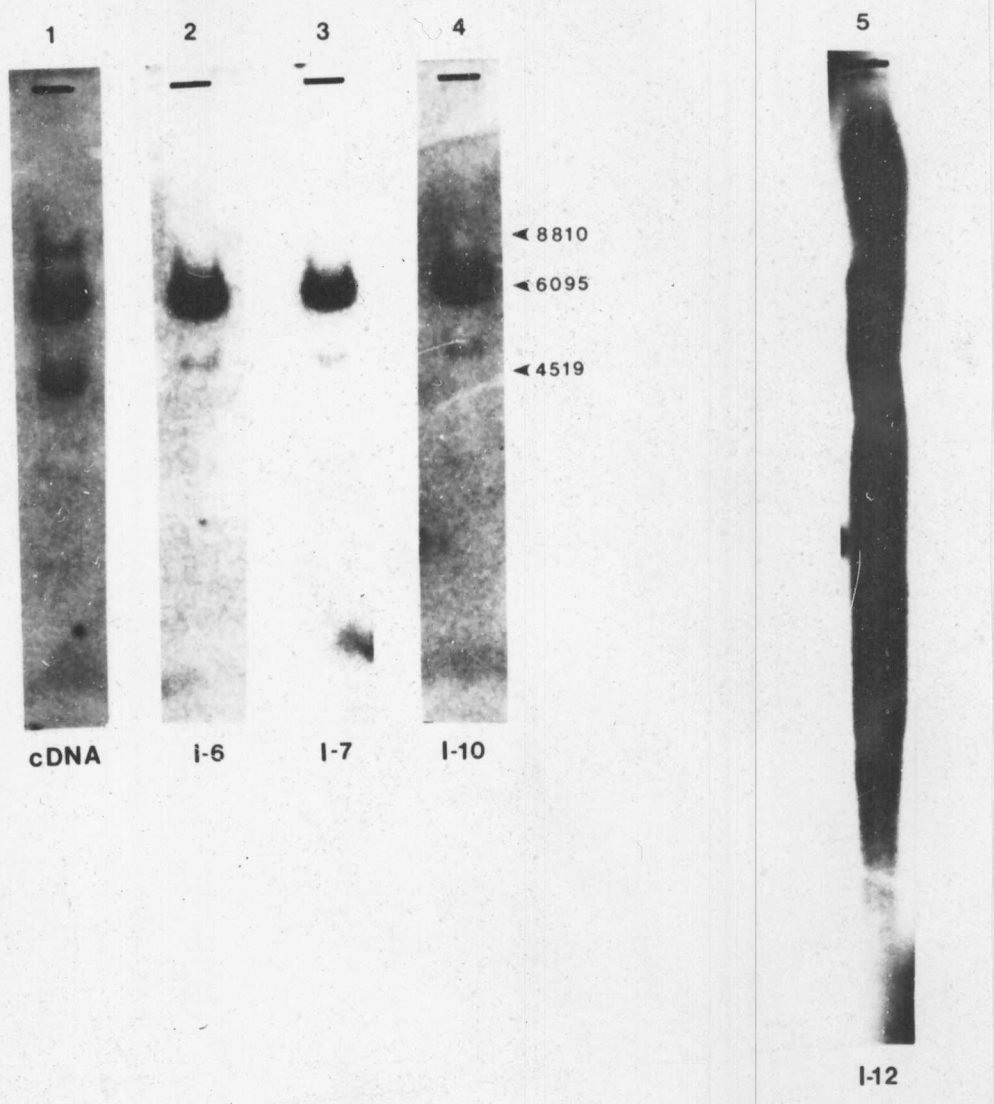


Figure III-6. Southern analysis of Eco RI digested genomic DNA with cDNA and intron probes.

The three fragments detected in lane 1 by the cDNA represent the three Eco RI fragments in the natural albumin gene that have regions homologous to the pMA615 cDNA probe. Introns 6, 7, and 10 detect only the middle 6 kb fragment (contained in Eco RI subclone 3-2) (lanes 2, 3, and 4). Lane 5 shows that intron 12 detects repeated sequences throughout the mouse genome.

characteristic of the hybridization expected for highly repetitive sequences.

The insert in the intron 12 subclone was known to contain two fragments of DNA, separated by an Sst I site, which might be useful in further localizing repeated sequences in this region. Several other intron 12 region fragments also remained uncharacterized in the 1-3 clone. Therefore, an experiment was performed to analyze the 1-12 region in greater detail. When nick translated, total genomic DNA is hybridized to albumin gene DNA fragments, hybridization is expected only with fragments that contain repeated sequences. This result is seen because unique sequences in the genome are in such low proportion to the whole DNA complement that they give an undetectable signal when hybridized (79,171).

When the 1-3 Eco RI subclone which contains the intron 12 region (see Figure III-4, page 47) is digested and hybridized with labelled genomic DNA, the results seen in Figure III-5C are seen. This experiment reveals that repeated sequences are present in both of the fragments in the intron 12 subclone and also in the adjacent 875 bp Bam HI fragment. This Bam HI fragment also contains exon "1" (see Figures III-5B and III-4) but the repeated sequence must not overlap the exon region because hybridization to genomic DNA with pMA615 (exon probe) detects only unique fragments (see Figure III-6). A control for the detection only of repeat sequences by the genomic probe is the fact that other unique sequence fragments do not show any signal with this probe (see the 680 bp and 550 bp bands in lane 4 of Figure III-5).

IV. DISCUSSION

The results presented here have added to our knowledge of the fine structure of the albumin gene. For all of the restriction enzyme sites originally indicated on a large scale map (88), approximate nucleotide distances have been established. In addition, some of the enzyme sites for Msp I have also been determined. When compared to the large scale albumin map of Kiousis et al. (89), differences in the locations of several Hind III restriction enzyme sites have been found. These differences could result from polymorphisms in this region of the gene or from errors in mapping. Another interesting possibility is a rearrangement of this region in association with an oncogenic transformation event. The origin of the sequences examined here was DNA from a plasmacytoma cell line, while those characterized by Kiousis et al. were from fetal mouse DNA (89). Gene rearrangements in oncogenesis have been described (135) but the limited nature of the differences seen here makes the alternate explanations more likely. Sequence analyses should resolve these differences in the near future. This new and more detailed information allows the examination of the albumin gene with the aim of identifying regions that might be useful in further functional analyses.

Four regions of intron DNA were determined to be bounded by restriction enzyme sites that could readily be matched to cloning sites in pBR322. These regions from introns 6, 7, 10, and 12 were, therefore, selected for subcloning. The characterization of the

intron subclones presented here shows that they do contain albumin specific regions because they identify the correct albumin fragment when hybridized to genomic DNA and that three of them represent sequences that are unique within the genome. With a cDNA probe containing exon regions available, the preparation of probes specific for introns 6, 7, and 10 provide the opportunity for performing several types of analyses comparing expressed with nonexpressed gene sequences. At the DNA level, these differential probes can be used to study differences in chromatin conformation or interactions in nuclear matrix associated functions. Also, experiments designed to follow the fate of intron regions during RNA processing have been carried out and are presented in Chapter IV. Note that the 5' Hind III sites of the intron 6 and 10 subclones lie at the edges of exon regions (see Figure III-2, page 45). Though at least a few of the subcloned nucleotides originate as exons, the Hind III sites must be close enough to the intron/exon boundary to prevent detectable hybridization with a cDNA probe.

A fourth subclone containing an intron 12 region has been shown to contain sequences that are highly repeated in the mouse genome. This fact makes hybridization studies such as those suggested above difficult, if not impossible, to interpret. However, this region will be useful in analyzing both the possible functional roles and the evolutionary origins of this type of sequence. If the repeated sequence does not enter the region of exon 1 (see Figure III-4, page 47), the size of this repeat must fall between 300 bp

and 1200 bp, as judged from the detected restriction fragments. This result most closely assigns this region to the group of short, interspersed repeats seen in many eukaryotes, including yeast (172), and agrees well with a size of about 925 bp determined by Kiousis et al. (88).

It may be speculated that the evolutionary origin of this region in an intron of albumin occurred by some type of transpositional mechanism. There is persuasive circumstantial evidence to support this notion. The corresponding intron of AFP does not have this sequence but its intron 3 does have a shorter related region as well as two homologous sequences in its 5' flanking region. Intron 12 is the central intron in a proposed four exon primordial domain (88) (see Figure II-1, page 15) and this repeat is not found in an analogous position in any of the other domain-centered introns of either albumin or AFP. In addition, despite the presence of many types of repeats known to be interspersed between genes (172), relatively few are found within introns (41,127). Rat growth hormone but not its human counterpart is an exception (127), and the proposal of insertion has been suggested to account for this observation also. Rat albumin intron 12 must also be free of this repeated sequence because when genomic DNA is used in hybridizations against clones of genomic rat DNA, only unique fragments are detected in the region of intron 12 (unpublished). These observations make it extremely unlikely that this repeated region was present either very early when the multiple-domain primordial gene formed (123) or at later

stages of evolution when the albumin and AFP duplication occurred (79) or when rat and mouse diverged. Instead, the proposal that these regions are somehow mobile in the genome and were inserted at a more recent time, better explains the observed distribution. Recently, other experiments have also detected homology of this region with a sequence derived from the mouse β-globin gene cluster (67). Another unrelated repeated sequence has been discovered in intron 1 of AFP and has also been suggested to be an inserted sequence (196). It is a member of the common mouse Alu-like family and possesses the flanking direct repeats and other features characteristic of sequences that have been proposed to have an insertion origin (196).

There is very little evidence for a function for these repeat regions as yet. A general observation that might argue for a function is that mRNA contains significant amounts of these sequences but at least 90% is lost during processing in the nucleus prior to transport into the cytoplasm (49). In the case of Alu-like repeats, significant homology has been found with the 7S RNA that is involved in protein secretion through the signal recognition particle (189). A reverse transcription of this RNA and then insertion into the genome has been proposed (81) but even if true, the reasons for selection of these insertion sites remains a mystery.

CHAPTER IV

ALBUMIN PRECURSOR RNAs ARE PROCESSED THROUGH VARIABLE PATHWAYS IN ASSOCIATION WITH THE NUCLEAR MATRIX

I. INTRODUCTION

The processing of precursor RNAs offer the eukaryotic cell another level of regulation of gene expression. Many genes, including albumin and AFP (183), have been shown to be regulated primarily at the level of gene transcription but, in recent years, post-transcriptional mechanisms of regulation have also been seen in several systems. Immunoglobulin heavy chains, of several classes, have now been shown to be synthesized from two types of mRNAs which specify either a membrane-bound or a secreted form of the protein (101). These two mRNAs are transcribed from the same gene and then differential pathways of precursor splicing regulate which form of mRNA is generated for export to the cytoplasm for translation. In this example, the regulatory decision is made in selecting alternate polyadenylation and splice sites at the 3'-end of the RNA molecule. The choice of polyadenylation site as a regulatory event for adenovirus (200) has been described in Chapter II. After the choice of polyadenylation sites has been made, the resulting RNAs are differentially processed to yield still further diversity in the final mRNA population (116). Alternate splice sites also regulate 3'-end changes in mRNAs coding for either calcitonin or a neuropeptide in

a tissue specific manner (6). For calcitonin (6) and α-amylase (197) variability in 5'-end mRNA sequences has also been noted, but in these cases the protein coding region are not altered. In the cases where the splicing pattern changes the expression of exons at the 3'-ends of the genes, these processing variations essentially represent choices in splicing pathways open to particular mRNA precursors. The nuclear factors which influence these decisions remain unknown.

In some cases, examination of nuclear RNA processing intermediates has revealed molecules created by multistep processing of specific introns. At least two steps have been reported to be required for the removal of the large intron from the mouse β-globin transcript (87) and an intron in the chick α_2-collagen RNA precursor (9). These multistep splicing reactions potentially offer another process that might be exploited in regulation by splicing.

The splicing pathways for several high abundance mRNAs have been studied using two approaches. For vitellogenin, ovalbumin, and ovomucoid, duplex molecules formed by hybridizing nuclear RNAs to genomic DNA clones were analyzed with the electron microscope (36, 158,185). This method has the advantage of allowing identification of specific introns that have been removed from individual splicing intermediates. These analyses showed that individual introns were removed, not in a random or obligatory order, but in a preferred order (36,158,185). No partially excised introns were reported in these studies. Another approach that has been used to study processing pathways is one in which intron specific probes have been

hybridized to nuclear RNA in an attempt to determine which of the precursor bands, normally seen with a cDNA probe, are no longer detected. When this analysis was applied to the same ovomucoid and ovalbumin precursors studied above with electron microscopy, the results agreed completely and allowed assignment of a preferred order of processing for the seven introns involved (36,185). This type of experiment has also been successfully applied to the analysis of immunoglobin kappa light chain RNA precursors (137).

Several groups have reported experiments suggesting that the processing of mRNA precursors occurs in association with a specific structural component of the nucleus. Pederson and associates have proposed that nascent RNAs are packaged by a characteristic set of proteins as they are transcribed (133) and, further, that the resulting heterogeneous nuclear ribonucleoprotein (hnRNP) particles can be isolated and shown to contain RNA precursor molecules for β-globin (134). These particles are isolated from nuclei disrupted by sonication (133) and therefore may afford only a partial glimpse of the in vivo functioning nuclear structure. Another method of nuclear fractionation has recently been reported that preserves the ultrastructural integrity of the nucleus. This nuclear matrix structure, seen in a variety of eukaryotic cells, is derived from nuclei that have been exposed to a detergent and were extracted in high salt solutions (17). Many of the nonhistone and most of the histone protein components of chromatin are lost during this procedure but, for adenovirus (103), β-globin (154) and chicken oviduct proteins

(39), precursor RNA molecules are reported to remain associated. In the oviduct experiments, the complete set of both ovomucoid and ovalbumin RNA precursors were associated exclusively with the matrix structure, as opposed to any soluble fraction, and provided evidence that supports the concept that processing may occur on the matrix (39). In support of this concept, the precursor rRNAs were also localized on the matrix structure while the U1 RNA, which has been implicated in RNA processing, was distributed between the matrix and soluble fractions (39). This function for nuclear matrix adds to those previously described which involved the DNA replication and transcription functions (16,150). Recently, a role of nuclear matrix in the export of RNP particles has been described (79).

In this chapter, evidence is presented that processing of albumin RNA precursors also occurs in a preferred order and that the precursor intermediates are associated with a residual nuclear protein matrix.

II. EXPERIMENTAL PROCEDURES

Analysis of Nuclear RNA Blots with Intron Probes

Nuclear RNA was prepared, blotted, and hybridized as described in Chapter I. Probes for introns 6, 7, and 10 were described in Chapter II.

Analysis of Nuclear Matrix Associated Albumin RNA

Nuclear matrix fractions were isolated using a modification of the procedure described by Ciejek et al. (39). Three livers were

excised into ice cold Buffer A (50% v/v glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.6) and homogenized in a tissue-mizer (Brinkman Instruments) at 0°C three times for 15 seconds each. The homogenate was filtered through ten layers of gauze into a 50 ml tube held at -20°C in a salt/ice mix. Crude nuclei were pelleted at 7500 rpm for 15 minutes at -20°C. The nuclear pellet was resuspended in 15 ml of ice cold Buffer B (2.2 M sucrose, 50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$) with three strokes in a Potter-Elvehjem type homogenizer. This solution was transferred to a precooled SW27 nitrocellulose tube and additional Buffer B added to fill the tube. Nuclei were pelleted at 55,000 xg for 75 minutes at 0°C in the SW27 rotor. Isolated nuclei were resuspended in solution A at -20°C with three strokes of a Potter-Elvehjem homogenizer and pelleted at 4000 xg for 10 minutes at -20°C. Nuclei were resuspended by aspiration in solution A plus 0.5% Triton X-100 and then pelleted as above. Nuclei were resuspended by aspiration in 2 ml of solution A, one-half of the preparation was held at -20°C for extraction of total nuclear RNA. To the other 1 ml was added 25 µl of a 4.2 mg/ml stock solution of RNase-free DNase I and this suspension was held for 5 minutes at zero degrees. The small degree of DNase activity under these conditions was determined empirically to be just enough to allow pelleting of nuclear matrices out of the gel that results when high salt is added to the subsequent step. To the digested nuclei was added 3 ml of solution C (2 M NaCl, 50% v/v glycerol, 1.5 mM $MgCl_2$, and 10 mM Tris-HCl, pH 7.4) at -20°C and this suspension was pelleted

at 4000 xg for 10 minutes. This pellet and the whole nuclear fraction above was lysed, phenol extracted, and the RNA purified through CsCl gradients as described for whole nuclei in Chapter II.

No commercially available preparation of DNase I is completely RNase-free, therefore DPFF grade DNase I (Worthington Biochemicals) was further purified by chromatography on a column of UMP-agarose (Miles-Veda Ltd.). The resulting solution was then adsorbed with bentonite (Fisher Chemicals) before use (64).

III. RESULTS

Analysis of Albumin Nuclear RNA Processing Pathways

Preliminary blotting experiments presented in Chapter II have shown reproducible albumin RNA processing intermediates from adult liver nuclei. In order to determine whether there is a specific sequence of processing, several unique sequence intron probes were isolated and characterized (see Chapter III). Four identical lanes of nuclear RNA were blotted onto nitrocellulose filters and were hybridized with either cDNA for albumin or one of the three sub-cloned intron regions. Figure IV-1 shows the autoradiograms of the detected RNA sequences. Lane 1 is the normal result that has been reproducibly seen with the albumin cDNA probe and shows a strong mRNA band. In these experiments a large hybridization signal in the area of the 45S rRNA is seen at the top of the lanes. This signal is seen occasionally in these nuclear RNA blot experiments and is believed to result from contaminating large molecular weight DNA for

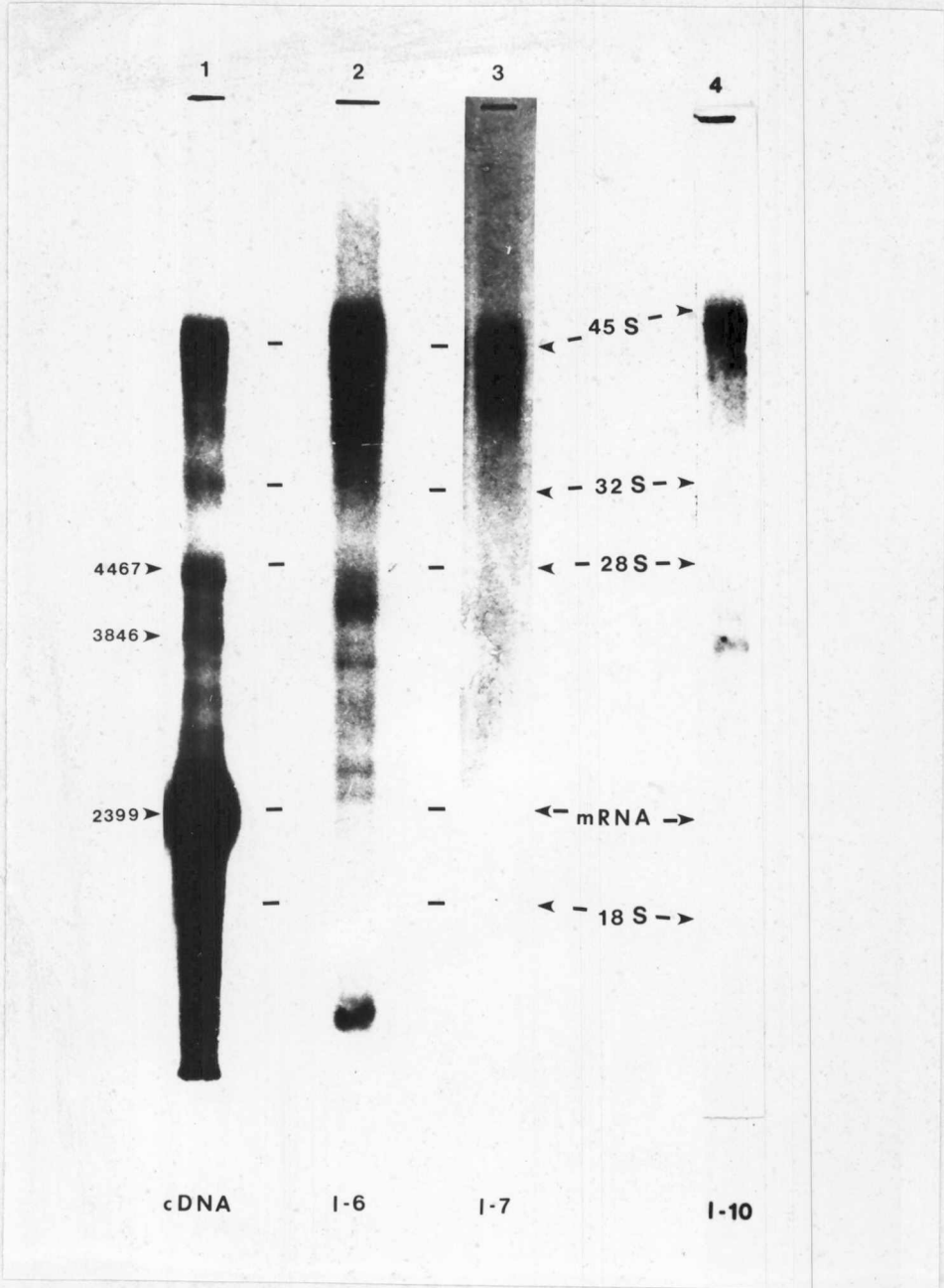


Figure IV-1. Analysis of albumin RNA processing pathway with intron subclones.

Ribosomal RNA standards and mature mRNA are indicated. Probes used in individual hybridization of each lane, containing 20 μg of nuclear RNA, are indicated below the lanes.

two reasons. The RNA preparation used in this experiment and several others that gave similar results gave a yield that was several fold greater than that usually seen. Also, a previously published experiment using this technique showed a very similar effect that was attributed to DNA contamination (72). It is believed that this contaminant does not affect the results shown in the lower portions of the lanes because the results seen with the cDNA are unchanged from those seen with this probe when no DNA signal is observed (see Figure II-2, page 21). Lanes 2-4 show the RNA precursors detected by individual intron probes and clearly show that the large amount of mRNA in these lanes is not detected by these probes. Intron 7 (lane 3) is shown to detect none of the usual RNA precursor bands suggesting that this intron is excised very early from the population of precursors. Intron 10 (lane 4) shows one band at about 3800 nt and several other faint ones. This result might be interpreted as showing a differential splicing pathway because several larger and more abundant intermediate species appear to have already lost this intron, yet, in some cases, smaller species do seem to retain it. Hybridization with intron 6 detects multiple species of RNAs but again does not hybridize to all of the same RNA precursors detected by the cDNA probe. This result also suggests a variable pathway of processing but, intron 6 in general, survives as a component of species that are smaller than those detected by intron 10 and some of which are very close to the mature mRNA size. Note that the intron 6 probe also strongly detects a small RNA species which is about

1300 nt in length similar in size to one that had been detected with the albumin cDNA probe.

Albumin RNA Processing Intermediates are Associated with Nuclear Matrix

Since the nuclear matrix had been implicated as an important multifunctional structure in the nucleus and had been proposed to be associated with precursor RNAs for several genes, nuclear matrix from adult mouse liver was analyzed for content of albumin RNA precursors. RNA isolated from a matrix preparation was separated on an agarose gel, blotted as described above, and hybridized with the labelled albumin cDNA probe. Figure IV-2 shows an experiment with nuclear matrix RNA (lane 2) and whole detergent washed nuclei (lane 1) that had not been DNase digested or extracted with high salt. The A and B parts of this figure show short and long exposures of the same experiment. Bands larger than the mature message are detectable in both of these lanes although large bands are difficult to resolve. The nuclear lane in this experiment contained approximately twice the amount of RNA as the matrix lane. The two small albumin bands that were previously seen are also detected associated with the matrix in this experiment. Previously reported RNA associations with matrix would suggest that these small albumin species might be involved in a processing pathway or are destined for cytoplasmic transport.

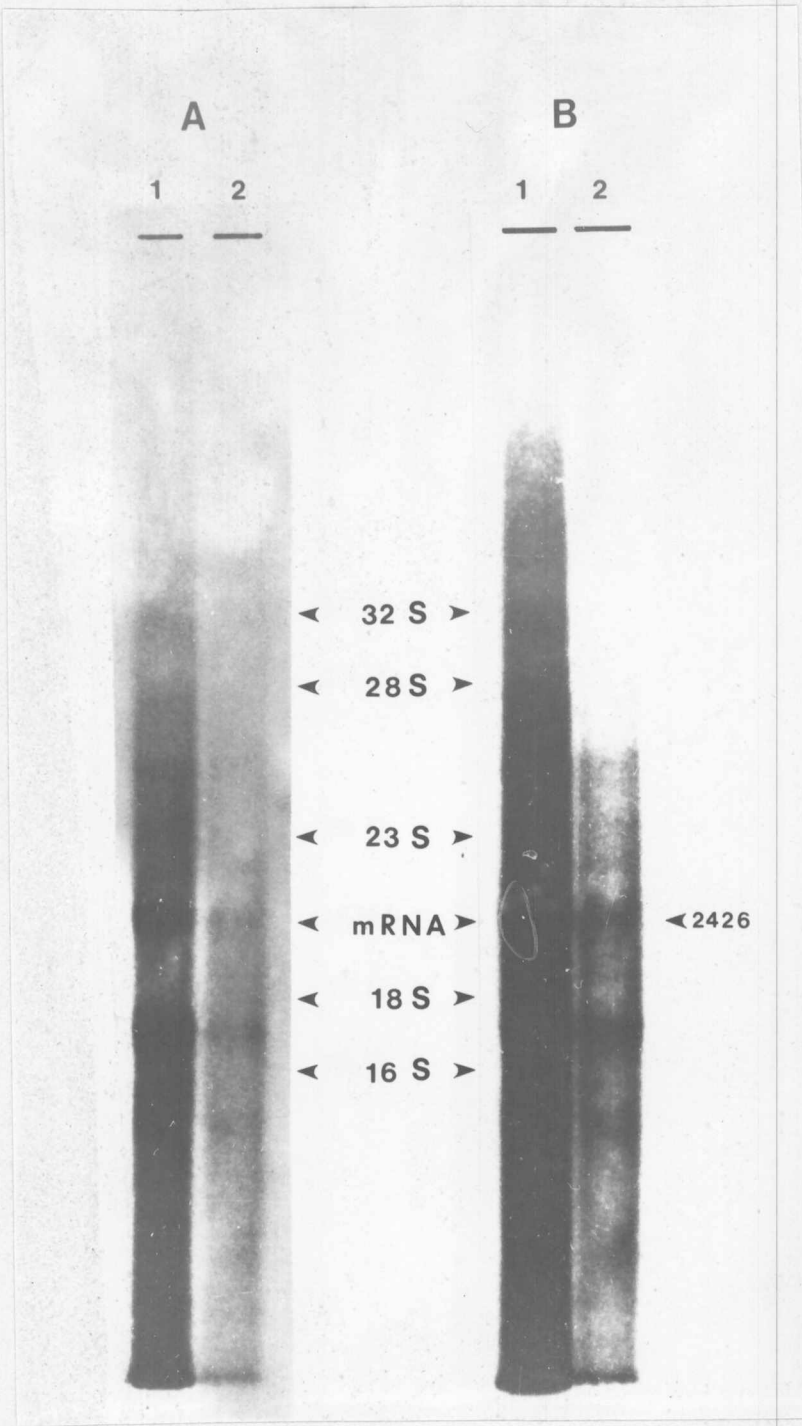


Figure IV-2. Albumin RNA processing intermediates associated with the nuclear protein matrix.

A and B parts are short (1.5 days) and long (7 days) autoradiogram exposures of the same experiment. Lane 1 contains about 32 μg of total nuclear RNA and lane 2 contains about 16 μg of nuclear matrix-bound RNA. Both lanes were transferred to filters and hybridized with the pMA615 probe.

IV. DISCUSSION

In general, a preferred order of removal of intron regions has been the finding in all previously studied systems of protein mRNA precursors. In some cases, a required order could be assigned to individual introns. For ovomucoid, intron E was not detected in any other species except the primary transcript (185) and in adenovirus, the first intron of the tripartite leader sequence seems to be removed first (116). This scenario might apply to intron 7 of albumin which is not detected in any nuclear RNA, but a firm conclusion cannot be drawn on this point because there must be early steps of processing that occur too rapidly to be detected in these experiments. A variable order of processing does seem consistent with the results seen for introns 10 and 6 probes. The existence of larger species of RNA precursors that are undetectable with these intron probes is good evidence that more than one processing pathway is available. Introns 6 and 10 are in positions in the albumin gene that separate the exon groups comprising the primitive domains (88) (see Figure II-1, page 15). One might expect some similarities in their processing fates because of this relationship but this similarity if it exists is not interpretable from these data.

The small RNA at about 1300 nt detected by intron 6 is very close to the predicted size of the intact intron 6 region of 1380 $\pm$ 93 bp (88). The most obvious interpretation of this result is that this intron is usually removed as an intact fragment that exists in a stable detectable concentration in the nucleus but does not

represent a functional sequence. In support of this idea, it has been found that this intron specific fragment is not polyadenylated and is not detectable in cytoplasmic RNA (data not shown). However, if the size of the smallest RNA precursor species (i.e., smallest band that is larger than the mRNA) that hybridizes to intron 6 is examined, it is found to be only several hundred nucleotides larger than the mRNA (see Figure IV-1, lane 2, page 64). Since intron 6 is 1380 nt long (88), there are two possible interpretations of this result. First is that this intron is removed by a multistep processing reaction thus generating the unexpected size intermediates. This interpretation cannot be ruled out at this time, but it may be relevant that in previous studies none of the high abundance mRNAs for proteins such as the oviduct proteins (185) or vitellogenin (158) showed multistep processing of individual introns. In addition, the observation of the strongly hybridizing band at 1300 nt (discussed above) suggests that most of this intron may be removed as one fragment. Another possible interpretation that might help explain the previously mentioned observations of small albumin RNA species (i.e., about 1800 nt and 1300 nt) is one of alternate splicing. The splicing process is known to exhibit flexibility in other systems so perhaps some unorthodox combinations of exons and introns are being synthesized by this process. An alternate splicing pathway might explain both the observation of unexpected sizes of RNA intermediates detected by intron 6 and the observation of small RNAs that contains sequences complementary to exon regions. This idea might explain

why these small species are also seen to be associated with nuclear matrix (Figure IV-2) and in cytoplasmic RNA (data not shown). In the nucleus (Figure II-2, page 21) and in the cytoplasm (data not shown) the small albumin RNA species seem to be distributed between poly[A-] populations (see Figure II-3, page 25) possibly indicative of short poly[A] tails. It is believed that the 1300 nt species detected in the nucleus by the intron 6 probe is distinct from the band of similar size detected by the cDNA probe because the hybridization signal detected in the exon experiments is usually clear but not strong while the band detected with the intron 6 probe (Figure IV-1, lane 2, page 64) is always quite intense. In addition, the two small bands detected with the exon probe are found in poly[A+] and cytoplasmic fractions while the 1300 nt intron 6 band is not found in these fractions. If alternate splicing occurs the question arises whether these small albumin RNAs are a result of a processing pathway that leads to a dead end or one that produces a functional molecule. If the small RNAs are nonfunctional, they seem to unexpectedly survive to reach the cytoplasm. The nuclear and cytoplasmic small albumin RNAs seen here are also seen in normal rat liver nuclei (56) but no interpretation is given for the presence of these RNAs in normal liver. Our data show a shift of at least several hundred nucleotides, from a 2200 nt mRNA to 1800 nt and another at 1300 nt (see Figure II-2). Therefore loss of a 200 nt poly[A] tail cannot account for this observation. One other possible origin for the small albumin RNA species is transcription from

the opposite DNA strand. Because the probes used in these experiments are labelled in both strands, an RNA of this type could be detected. A precedent for transcription on the opposite strand of AFP has been shown to occur at an initiation site in a repeated sequence in intron 1 (196). A repeated sequence, possibly as seen in intron 12 of the albumin gene, might be the most likely promoter for this type of transcription. Cellular turnover of mRNA occurs naturally and during this process secondary or tertiary structural features of the mRNP particle may contribute to the detection of discrete bands of degraded RNA. Therefore, even though precautions have been taken, the possibility cannot be excluded that the small albumin RNA bands result from a specific degradation of mature mRNA. If these small RNAs do somehow function, then they represent a fascinating new feature of albumin gene expression and one that should be amenable to deciphering with current techniques.

In these nuclear RNA experiments, it has always been difficult to predict with confidence the preparation of undegraded RNA. Originally, it was hoped that nuclear matrix preparations would provide an RNA fraction that was enriched for precursors as would be expected if bulk nonprecursor RNA was removed from the matrix fraction. Unfortunately, preparations of nuclear matrix RNA were found to be even more difficult to maintain in an undegraded state. The experiment shown in Figure IV-2, page 67, does show detectable albumin RNA species that are larger than the messenger RNA but the normal pattern of albumin precursors is not seen. The fading array

of precursors, seen in Figure IV-2, page 67, coupled with the fact that in this gel the 45S rRNA precursor could not be detected (data not shown), suggests that some degradation has occurred in these samples. However, the conclusion that at least some of the precursor molecules and the small albumin RNA molecules are matrix associated seems reasonable. Though the data for matrix association are not as complete as was hoped when these experiments were begun, they seem consistent with an emerging unifying concept of a structural framework in eukaryotic nuclei. If one considers the variety and complexity of cytoplasmic structures devoted to the synthesis, modification, processing, and transport of protein molecules, some type of analogous functional framework in the nucleus might be expected. The data contributed by many systems suggest that genes that are to undergo transcription become preferentially associated with the matrix structure (150), are transcribed and their transcripts are packaged into hnRNP particles and are then processed as they proceed in some unknown fashion towards the nuclear pore.

The contributions that this work on albumin and AFP nuclear RNA has made to molecular biology can be expressed in two parts. Experiments that have demonstrated multiple species of polyadenylated mRNA precursors, whose introns are removed in a preferred order, essentially confirm, in this newly studied system, earlier findings. The complex organization of the albumin and AFP genes would make further studies of the details of their RNA processing very difficult. Therefore, simpler systems of processing extracts studied in

vitro (126) or studies of "minigenes" constructed *in vitro* (196) are the proposed directions for future research.

The second class of presented results and their interpretations are of a more hypothetical nature. The notion that a functional small albumin RNA molecule might be generated from albumin RNA precursors remains only a preliminary proposal at present. Though precedents for regulation of this type do exist (5,6) at least two facts must be experimentally established before this suggestion can be seriously considered. First, the small RNAs must be shown not to be derived from a process of general degradation, or conversely that they do derive from a specific processing pathway. Second, a function for these small albumin RNA species should be confirmed. Experiments designed to determine what gene sequences are contained in these RNA molecules or what protein products might result from their translation should have relevance to these questions. Tissue specific variation in rate limiting steps of albumin RNA processing has also been observed here. The presence of this variation in the liver of newborns and tumor tissues most likely reflects some differences in the assembly of hnRNP particles for albumin transcripts. A different set of hnRNP proteins may have as its direct consequence changes in the set of detectable albumin RNA intermediates and at the same time be indicative of another method used by the genetic apparatus to establish a particular cellular phenotype. Gene expression differences in different tissues are often accounted for by transcription of different sets of genes, however, regulation by

alternate hnRNP assembly schemes could also affect RNA processing, translational efficiency and mRNA stability. Thus, hnRNP assembly may represent a cellular mechanism of fine tuning the expression of a particular assortment of transcribed genes. Characterization of tissue specific sets of nuclear proteins is possible using current technology, however, identifying their associations with particular nuclear transcripts will be the challenge of future research in this area.

REFERENCES

REFERENCES

1. Abelev, G. I. (1971). Alpha-fetoprotein in oncogenesis and its association with malignant tumors. Ad. Canc. Res. 14:295-357.

2. Abelev, G. I. Experimental study of alpha-fetoprotein re-expression in liver regeneration and hepatocellular carcinomas, in Cell Differentiation and Neoplasia. Grady F. Saunders ed. (New York: Raven Press, 1978), pp. 257-269.

3. Adesnik, M., Salditt, M., Thomas, W., and Darnell, J. E. (1972). Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly(A) sequences and that the poly(A) has a nuclear function. J. Mol. Biol. 71:21-30.

4. Adolph, K. W., Cheng, S. M., and Laemmli, U. K. (1977). Role of nonhistone proteins in metaphase chromosome structure. Cell 12:805-816.

5. Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin Mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell 20:293-301.

6. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S., and Evans, R. M. (1982). Alternative RNA processing in calctonin gene expression generates mRNAs encoding different polypeptide products. Nature 298:240-244.

7. Artymiuk, P. J., Blake, C. C. F., and Sippel, A. E. (1981). Genes pieced together--exons delineate homologous structures of diverged lysozymes. Nature 290:287-288.

8. Aussel, C., Stora, C., and Krebs, B. (1980). Alpha-fetoprotein and serum hormone levels following liver intoxication with carbon tetrachloride. Biochem. Biophys. Res. Comm. 95:796-800.

9. Avvidimento, V. E., Vogeli, G., Yamada, Y., Maizel, J. V., Pastan, I., and de Crombrugghe, B. (1980). Correlation between splicing sites within an intron and their sequence complementarity with U1 DNA. Cell 21:689-696.

10. Baglia, F. A., and Maul, G. G. (1983). Nuclear ribonucleoprotein release and nucleoside triphosphatase activity are inhibited by antibodies directed against one nuclear matrix glycoprotein. Proc. Natl. Acad. Sci. 80:2285-2289.

11. Basler, J., Hastie, N. D., Pietras, D., Matsui, S.-I., Sandberg, A. A., and Berezney, R. (1981). Hybridization of nuclear matrix attached deoxyribonucleic acid. Biochemistry 20:6921-6929.

12. Bastos, R. N., and Aviv, H. (1977). Globin RNA precursor molecules: Biosynthesis and processing in erythroid cells. Cell 11:641-650.

13. Bélanger, L., Hamel, D., Dufour, D., Guillouzo, A., and Chiu, J. F. (1978). Hormonal control and putative cell cycle frequency of AFP production: Further observations in vivo and in vitro. Scand. J. Immunol. 8:Suppl. 8:239-246.

14. Belayew, A., and Tilghman, S. M. (1982). Genetic analysis of α-fetoprotein synthesis in mice. Mol. Cell. Biol. 2:1427-1435.

15. Benyajati, C., and Worcel, A. (1976). Isolation, characterization and structure of the folded interphase genome of drosophila melanogaster. Cell 9:393-407.

16. Berezney, R., and Buchholtz, L. A. (1981). Dynamic association of replicating DNA fragments with the nuclear matrix of regenerating liver. Expt. Cell Res. 132:1-13.

17. Berezney, R., and Coffey, D. S. (1974). Identification of a nuclear protein matrix. Biochem. Biophys. Res. Commun. 60: 1410-1470.

18. Berezney, R., and Coffey, D. S. (1975). Nuclear protein matrix: Association with newly synthesized DNA. Science 189:291-293.

19. Berezney, R., and Coffey, D. S. (1976). The nuclear protein matrix: Isolation, structure, and functions. Adv. Enz. Regul. 14:63-100.

20. Berezney, R., and Coffey, D. S. (1977). Nuclear matrix: Isolation and characterization of a framework structure from rat liver nuclei. J. Cell Biol. 73:616-637.

21. Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. 74:3171-3175.

22. Bina, M., Feldmann, R. J., and Deeley, R. G. (1980). Could poly(A) align the splicing sites of messenger RNA precursors? Proc. Natl. Acad. Sci. 77:1278-1282.

23. Bingham, P. M., Kidwell, M. G., and Rubin, G. M. (1982). The molecular basis of P-M hybrid dysgenesis: The role of the P element, a P-strain-specific transposon family. Cell 29:995-1004.

24. Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7:1513-1523.

25. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Fiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. (1977). Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. Science 196:161-169.

26. Bowen, B. C. (1981). DNA fragments associated with chromosome scaffolds. Nucl. Acids Res. 9:5093-5108.

27. Brack, C., and Tonegawa, S. (1977). Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. Proc. Natl. Acad. Sci. 74: 5652-5656.

28. Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50:349-383.

29. Breathnach, R., Mandel, J. L., and Chambon, P. (1977). Ovalbumin gene is split in chicken DNA. Nature 270:314-319.

30. Brown, J. R. (1976). Structural origins of mammalian albumin. Fed. Proc. 35:2141-2144.

31. Brown, P. C., and Papaconstantinou, J. (1977). Identification of albumin-synthesizing polysomes from mouse liver and a mouse hepatoma cell line. Biochem. Biophys. Res. Commun. 76:121-127.

32. Brown, P. C., and Papaconstantinou, J. (1979). Mouse albumin mRNA in liver and a hepatoma cell line. J. Biol. Chem. 254:5177-5183.

33. Busch, H., and Daskal, Y. (1977). Methods for isolation of nuclei and nucleoli. Meth. Cell Biol. 16:1-43.

34. Capco, D. G., Wan, K. M., and Penman, S. (1982). The nuclear matrix: Three-dimensional architecture and protein composition. Cell 29:847-858.

35. Cech, T. R. (1983). RNA splicing: Three themes with variations. Cell 34:713-716.

36. Chambon, P., Benoist, C., Breathnach, R., Cochet, M., Gannon, F., Gerlinger, P., Krust, A., Le Meur, M., Le Pennec, J. P., Mandell, J. L., O'Hare, K., and Perrin, F. Structural organization and expression of ovalbumin and related chicken genes in From Gene to Protein: Information Transfer in Normal and Abnormal Cells, T. R. Russell, K. Brew, H. Faber, and J. Schultz, eds. (New York: Academic Press, Inc., 1979), pp. 85-98.

37. Chiu, J. F., Decha-Umphai, W., and Commer, P. (1979) α_1-Fetoprotein mRNA of rat yolk sac and hepatoma. Nucl. Acids Res. 7: 239-249.

38. Chu, G., and Sharp, P. A. (1981). A gene chimaera of SV40 and mouse β-globin is transcribed and properly spliced. Nature 289: 378-382.

39. Ciejek, E. M., Nordstrom, J. L., Tsai, M.-J., and O'Malley, B. W. (1982). Ribonucleic acid precursors are associated with the chick oviduct nuclear matrix. Biochemistry 21:4945-4953.

40. Clewell, D. B., and Helinski, D. R. (1970). Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. Biochemistry 9:4428-4440.

41. Cochet, M., Gannon, F., Hen, R., Maroteaux, F., Perrin, F., and Chambon, P. (1979). Organization and sequence studies of the 17-piece chicken conalbumin gene. Nature 282:567-574.

42. Cook, P. R., and Brazell, I. A. (1976). Conformational constraints in nuclear DNA. J. Cell Sci. 22:287-302.

43. Cook, P. R., and Brazell, I. A. (1980). Mapping sequences in loops of nuclear DNA by their progressive detachment from the nuclear cage. Nucl. Acids Res. 8:2895-2908.

44. Craik, C. S., Buchman, S. R., and Beychok, S. (1980). Characterization of globin domains: Heme binding to the central exon product. Proc. Natl. Acad. Sci. 77:1384-1388.

45. Curtis, P. J., and Weissman, C. (1976). Purification of globin messenger RNA from dimethylsulfoxide-induced Friend cells and detection of a putative globin messenger RNA precursor. J. Mol. Biol. 106:1061-1075.

46. Dagert, M., and Ehrlich, S. D. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6:23-28.

47. Darnell, J. E. (1978). Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science 202:1257-1260.

48. Darnell, J. E., Jelinek, W. R., and Molloy, G. R. (1973). Biogenesis of mRNA: Genetic regulation in mammalian cells. Science 181:1215-1221.

49. Davidson, E. H., and Britten, R. J. (1979). Regulation of gene expression: Possible role of repetitive sequences. Science 204:1052-1059.

50. De La Salle, H., Jacq, C., and Slonimski, P. P. (1982). Critical sequences within mitochondrial introns: Pleiotropic mRNA maturase and cis-dominant signals of the box intron controlling reductase and oxidase. Cell 28:721-732.

51. Doolittle, W. F. (1978). Genes in pieces: Were they ever together? Nature 272:581-582.

52. Doolittle, W. F., and Sapienza, C. (1983). Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601-603.

53. Dugaiczyk, A. C., Law, S. W., and Dennison, O. E. (1982). Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. Proc. Natl. Acad. Sci. 79:71-75.

54. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell 20:313-319.

55. Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981). Intragenic amplification and divergence in the mouse α-fetoprotein gene. Nature 294:713-718.

56. Esumi, H., Takahashi, Y., Sekiya, T., Sato, S., Nagase, S., and Sugimura, T. (1982). Presence of albumin mRNA precursors in nuclei of analbuminemic rat liver lacking cytoplasmic albumin mRNA. Proc. Natl. Acad. Sci. 79:734-738.

57. Federoff, N. V. (1979). On spacers. Cell 16:697-710.

58. Fraser, N. W., Nevins, J. R., Ziff, E., and Darnell, J. E. (1979). The major late adenovirus type-2 transcription unit: Termination is downstream from the last poly[A] site. J. Mol. Biol. 129:643-656.

59. Galli, G., Hofstetter, H., Stunnenberg, H. G., and Birnstiel, M. L. (1983). Biochemical complementation with RNA in the Xenopus oocyte: A small RNA is required for the generation of 3' histone mRNA termini. Cell 34:823-828.

60. Gilbert, W. (1978). Why genes in pieces? Nature 271:501.

61. Goldsmith, M. E., Humphries, R. K., Ley, T., Cline, A., Kantor, J. A., and Nienhuis, A. W. (1983). "Silent" nucleotide substitution in a β^+-thalassemia globin gene activates splice site in coding sequence RNA. Proc. Natl. Acad. Sci. 80:2318-2322.

62. Gorin, M. B., Cooper, D. L., Eiferman, F., van de Rign, P., and Tilghman, S. M. (1981). The evolution of α-fetoprotein and albumin. J. Biol. Chem. 256:1954-1959.

63. Gorin, M. B., and Tilghman, S. M. (1980). Structure of the α-fetoprotein gene in the mouse. Proc. Natl. Acad. Sci. 77:1351-1355.

64. Grady, L. J., Campbell, W. P., and North, A. B. (1980). A comparison of several procedures for reducing RNase contamination in preparations of DNase and a particularly successful combination of methods. Anal. Biochem. 101:118-122.

65. Green, M. R., Maniatis, T., and Melton, D. A. (1983). Human β-globin pre-mRNA synthesized *in vitro* is accurately spliced in *Xenopus* oocyte nuclei. Cell 32:681-694.

66. Gruss, P., Lai, C.-J., Dhar, R., and Khoury, G. (1979). Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proc. Natl. Acad. Sci. 76:4317-4321.

67. Haigwood, N. L., Jahn, C. L., Hutchison, C. A., and Edgell, M. H. (1981). Locations of three repetitive families found in Balb/c adult β-globin clones. Nucl. Acids Res. 9:1133-1150.

68. Hamer, D. H., and Leder, P. (1979). Splicing and the formation of stable RNA. Cell 18:1299-1302.

69. Harpold, M. M., Dobner, P. R., Evans, R., Bancroft, F. C., and Darnell, J. E. (1979). The synthesis and processing of a nuclear RNA precursor to rat pregrowth hormone messenger RNA. Nucl. Acids Res. 6:3133-3144.

70. Hastings, K. E. M., and Emerson, C. P. (1982). cDNA clone analysis of six co-regulated mRNAs encoding skeletal muscle contractile proteins. Proc. Natl. Acad. Sci. 79:1553-1557.

71. Hawkins, J. W., and Dugaiczyk, A. C. (1982). The human serum albumin gene: Structure of a unique locus. Gene 19:55-58.

72. Henikoff, S., Sloan, J. S., and Kelly, J. D. (1983). A Drosophila metabolic gene transcript is alternatively processed. Cell 35:405-414.

73. Herlan, G., Eckert, W. A., Kaffenberger, W., and Wunderlich, F. (1979). Isolation and characterization of an RNA-containing nuclear matrix from Tetrahymena macronuclei. Biochemistry 18: 1782-1788.

74. Hermann, R. C., Weymouth, L., and Penman, S. (1978). Heterogeneous nuclear RNA-protein fibers in chromatin-depleted nuclei. J. Cell Biol. 78:663-674.

75. Hodge, L. D., Mancini, P., Davis, F. M., and Heywood, P. (1977). Nuclear matrix of HeLa S_3 cells: Polypeptide composition during adenovirus infection and in phases of the cell cycle. J. Cell Biol. 72:194-208.

76. Hofer, E., and Darnell, J. E. (1981). The primary transcription unit of the mouse β-major globin gene. Cell 23:585-593.

77. Hofer, E., Hofer-Warbiner, R., and Darnell, J. E. (1982). Globin RNA transcription: A possible termination site and demonstration of transcriptional control correlated with altered chromatin structure. Cell 29:887-893.

78. Hoffman, L. M., Fritsch, M. K., and Gorski, J. (1981). Probable nuclear precursors of preprolactin mRNA in rat pituitary cells. J. Biol. Chem. 256:2597-2600.

79. Ingram, R. S., Scott, R. W., and Tilghman, S. M. (1981). α-Fetoprotein and albumin genes are in trandem in the mouse genome. Proc. Natl. Acad. Sci. 78:4694-4698.

80. Jackson, D. A., McCready, S. J., and Cook, P. R. (1981). RNA is synthesized at the nuclear cage. Nature 292:552-555.

81. Jagadeewaran, P., Forget, B. G., and Weissman, S. M. (1981). Short interspersed repetitive DNA elements in eucaryotes: Transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell 26:141-142.

82. Jagodzinki, L. L., Sargent, T. D., Yang, M., Glackin, C., and Bonner, J. (1981). Sequence homology between RNAs encoding rat α-fetoprotein and rat serum albumin. Proc. Natl. Acad. Sci. 78:3521-3525.

83. Jeffreys, A. J., and Flavell, R. A. (1977). The rabbit β-globin gene contains a large insert in the coding sequence. Cell 12:1097-1108.

84. Kaufmann, S. H., Coffey, D. S., and Shaper, J. H. (1981). Considerations in the isolation of rat liver nuclear matrix, nuclear envelope, and pore complex lamina. Expt. Cell Res. 132:105-123.

85. Keller, G. H., and Taylor, J. M. (1976). Effect of hypophysectomy on the synthesis of rat liver albumin. J. Biol. Chem. 251:3768-3773.

86. Kinniburgh, A. J., Mertz, J. E., and Ross, J. (1978). The precursor of a mouse β-globin messenger RNA contains two intervening RNA sequences. Cell 14:681-693.

87. Kinniburgh, A. J., and Ross, J. (1979). Processing of the mouse β-globin mRNA precursor: At least two cleavage-ligation reactions are necessary to excise the larger intervening sequence. Cell 17:915-921.

88. Kioussis, D., Eiferman, F., van de Rijn, P., Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1981). The evolution of α-fetoprotein and albumin. J. Biol. Chem. 256:1960-1967.

89. Kleckner, N. (1977). Translocatable elements in procaryotes. Cell 11:11-23.

90. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A., and Georgiev, G. P. (1980). The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucl. Acids Res. 8:1201-1215.

91. Kwan, S.-P., Wood, T. G., and Lingrel, J. B. (1977). Purification of a putative precursor of globin messenger RNA from mouse nucleated erythroid cells. Proc. Natl. Acad. Sci. 74:178-182.

92. Lang, K. M., and Spritz, R. A. (1983). RNA splice site selection: Evidence for a 5' → 3' scanning model. Science 220: 1351-1355.

93. Langford, C., and Gallwitz, D. (1983). Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell 33:519-527.

94. Langford, C., Nellen, W., Niessing, J., and Gallwitz, D. (1983). Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. Proc. Natl. Acad. Sci. 80:1496-1500.

95. Law, S. W., and Dugaiczyk, A. (1981). Homology between the primary structure of α-fetoprotein, deduced from a complete cDNA sequence, and serum albumin. Nature 291:201-205.

96. Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 16:4743-4751.

97. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980). Are snRNPs involved in splicing? Nature 283:220-224.

98. Levis, R., and Penman, S. (1977). The metabolism of poly(A)+ and poly(A)-hnRNA in cultured Drosophila cells studied with a rapid uridine pulse-chase. Cell 11:105-113.

99. Lewin, R. (1982). On the origin of introns. Science 217: 921-922.

100. Liao, W. S. L., Conn, A. R., and Taylor, J. M. (1980). Changes in rat α_1-fetoprotein and albumin mRNA levels during fetal and neonatal development. J. Biol. Chem. 255:10036-10039.

101. Liu, C.-P., Tucker, P. W., Mushinski, J. F., and Blattner, F. R. (1980). Mapping of heavy chain genes for mouse immunoglobulins M and D. Science 209:1348-1353.

102. Long, B. H., Huang, C.-Y., and Pogo, A. D. (1979). Isolation and characterization of the nuclear matrix in Friend erythroleukemia cells: Chromatin and hnRNA interactions with the nuclear matrix. Cell 18:1079-1090.

103. Mariman, E. C. M., van Eekelen, C. A. G., Reinders, R. J., Berns, A. J. M., and van Venrooij, W. J. (1982). Adenoviral heterogeneous nuclear RNA is associated with the host nuclear matrix during splicing. J. Mol. Biol. 154:103-119.

104. Marx, J. L. (1983). A transposable element of maize emerges. Science 219:829-830.

105. Maurer, R. A. (1980). Immunochemical isolation of prolactin messenger RNA. J. Biol. Chem. 255:854-859.

106. McCready, S. J., Akrigg, A., and Cook, P. R. (1979). Electron miscroscopy of intact nuclear DNA from human cells. J. Cell Sci. 39:53-62.

107. McCready, S. J., Godwin, J., Mason, D. W., Brazell, I. A., and Cook, P. R. (1980). DNA is replicated at the nuclear cage. J. Cell Sci. 46:365-386.

108. Milcarek, C. (1979). HeLa cell cytoplasmic mRNA contains three classes of sequences: Predominantly poly(A)-free, predominantly poly(A)-containing and bimorphic. Eur. J. Biochem. 102:467-476.

109. Miller, T. E., Huang, C.-Y., and Pogo, A. O. (1978). Rat liver nuclear skeleton and ribonucleoprotein complexes containing hnRNA. J. Cell Biol. 76:675-691.

110. Montell, C., Fisher, E. F., Caruthers, M. H., and Berk, A. J. (1983). Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. Nature 305:600-605.

111. Moringa, T., Sakai, M., Wegmann, T. G., and Tamaoki, T. (1983). Primary structures of human α-fetoprotein and its mRNA. Proc. Natl. Acad. Sci. 80:4606-4608.

112. Mount, S. M., and Steitz, J. (1983). Lessons from mutant globins. Nature 303:380-381.

113. Mulvihill, E. R., Le Pennec, J.-P., and Chambon, P. (1982). Chicken oviduct progesterone receptor: Location of specific regions of high-affinity binding in cloned DNA fragments of hormone-responsive genes. Cell 24:621-632.

114. Nelkin, B. D., Pardoll, D. M., and Vogelstein, B. (1980). Localization of SV40 genes within supercoiled loop domains. Nucl. Acids Res. 8:5623-5633.

115. Nevers, P., and Saedler, H. (1977). Transposable genetic elements as agents of gene instability and chromosomal rearrangements. Nature 268:109-115.

116. Nevins, J. R., and Chen-Kiang, S. (1981). Processing of adenovirus nuclear RNA to mRNA. Adv. in Virus Res. 26:1-35.

117. Nevins, J. R., and Darnell, J. E. (1978). Steps in the processing of Ad2 mRNA: Poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell 15:1477-1493.

118. Nevins, J. R., and Wilson, M. C. (1981). Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. Nature 290:113-118.

119. Nordstrom, J. L., Roop, D. R., Tsai, M.-J., and O'Malley, B. W. (1979). Identification of potential ovomucoid mRNA precursors in chick oviduct nuclei. Nature 278:328-331.

120. Norgard, M. W., Emigholz, K., and Monahan, J. J. (1979). Increased amplification of pBR322 plasmid deoxyribonucleic acid in Escherichia coli K-12 strains RR1 and χ1776 grown in the presence of high concentrations of nucleoside. J. Bact. 138: 270-272.

121. O'Farrell, P. Z., Cordell, B., Valenzuela, P., Rutter, W. J., and Goodman, H. M. (1978). Structure and processing of yeast precursor tRNAs containing intervening sequences. Nature 274:438-445.

122. Ohkubo, H., Vogelli, G., Mudryj, M., Avvedimento, E., Sullivan, M., Paston, I., and de Crumbugge, B. (1980). Isolation and characterization of overlapping genomic clones covering the chicken 2 (type I) collagen gene. Proc. Natl. Acad. Sci. 77: 7059-7063.

123. Ohno, S. (1981). Original domain for the serum albumin family arose from repeated sequences. Proc. Natl. Acad. Sci. 78: 7657-7661.

124. Ohshima, Y., Itoh, M., Okada, N., and Miyata, T. (1981). Novel models for RNA splicing that involve a small nuclear RNA. Proc. Natl. Acad. Sci. 78:4471-4474.

125. Orgel, L. E., and Crick, F. H. C. (1980). Selfish DNA: The ultimate parasite. Nature 284:604-607.

126. Padgett, R. A., Hardy, S. F., and Sharp, P. A. (1983). Splicing of adenovirus RNA in a cell-free transcription system. Proc. Natl. Acad. Sci. 80:5230-5234.

127. Page, G. S., Smith, S., and Goodman, H. M. (1981). DNA sequence of the rat growth hormone gene: Location of the 5' terminus of the growth hormone mRNA and identification of an internal transposon-like element. Nucl. Acids Res. 9:2087-2104.

128. Palmiter, R. D. (1974). Magnesium precipitation of ribonucleoprotein complexes. Expedient techniques for the isolation of undegraded polysomes and messenger ribonucleic acid. Biochemistry 13:3606-3615.

129. Pardoll, D. M., and Vogelstein, B. (1980). Sequence analysis of nuclear matrix associated DNA from rat liver. Exp. Cell Res. 128:466-470.

130. Pardoll, D. M., Vogelstein, B., and Coffey, D. S. (1980). A fixed site of DNA replication in eucaryotic cells. Cell 19: 527-536.

131. Parker, M. (1983). Enhancer elements activated by steroid hormones? Nature 304:687-688.

132. Paulson, J. R., and Laemmli, U. K. (1977). The structure of histone-depleted metaphase chromosomes. Cell 12:817-828.

133. Pederson, T. (1981). Messenger RNA biosynthesis and nuclear structure. American Scientist 69:76-84.

134. Pederson, T., and Davis, N. B. (1980). Messenger RNA processing and nuclear structure: Isolation of nuclear ribonucleoprotein particles containing β-globin messenger RNA precursors. J. Cell Biol. 87:47-54.

135. Perry, R. P. (1983). Consequences of myc invasion of immunoglobulin loci: Facts and speculation. Cell 33:647-649.

136. Perry, R. P., and Kelley, D. E. (1979). Immunoglobin messenger RNAs in murine cell lines that have characteristics of immature B lymphocytes. Cell 18:1333-1339.

137. Perry, R. P., Kelley, D. E., Coleclough, C., Seidman, J. G., Leder, P., Tonegawa, S., Matthyssens, G., and Weigert, M. (1980). Transcription of mouse κ chains. Proc. Natl. Acad. Sci. 77:1937-1941.

138. Perry, R. P., Kelley, D. E., and Schibler, U. (1979). Comparison of mRNA precursors in plasmacytomas producing closely related κ chains. Proc. Natl. Acad. Sci. 76:3678-3682.

139. Peters, T. (1977). Serum albumin: Recent progress in the understanding of its structure and biosynthesis. Clin. Chem. 23:5-12.

140. Peters, T., and Peters, J. C. (1972). The biosynthesis of rat serum albumin. J. Biol. Chem. 247:3858-3863.

141. Pikielney, C. W., Teem, J. L., and Rosbash, M. (1983). Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: Implication for U1 RNA and metazoan mRNA splicing. Cell 34:395-403.

142. Prentki, P., Karch, F., Iida, S., and Meyer, J. (1981). The plasmid cloning vector pBR325 contains a 482 base-pair-long inverted duplication. Gene 14:282-299.

143. Proudfoot, N. J., and Brownlee, G. G. (1976). 3' Noncoding region sequences in eukaryotic messenger RNA. Nature 263:211-214.

144. Razin, A., and Riggs, A. D. (1980). DNA methylation and gene function. Science 210:604-610.

145. Reanney, D. C. (1974). On the origin of prokaryotes. J. Theor. Biol. 48:243-251.

146. Revel, M., and Groner, Y. (1978). Post-transcriptional and translational controls of gene expression in eukaryotes. Ann. Rev. Biochem. 47:1079-1126.

147. Rigby, P. W. J., Dieckman, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113:237-251.

148. Robin, C., Hauck, M., Deininger, P. L., Friedmann, T., and Schmid, C. W. (1980). Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature 284:372-374.

149. Robinson, S. I., Nelkin, B. D., and Vogelstein, B. (1982). The ovalbumin gene is associated with the nuclear matrix of chicken oviduct cells. Cell 28:99-106.

150. Robinson, S. I., Small, D., Idyerda, R., McKnight, G. S., and Vogelstein, B. (1983). The association of transcriptionally active genes with the nuclear matrix of the chicken oviduct. Nucl. Acids Res. 11:5113-51130.

151. Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell 20:303-312.

152. Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. Proc. Natl. Acad. Sci. 77:1877-1879.

153. Roop, D. R., Nordstrom, J. L., Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. (1978). Transcription of structural and intervening sequences in the ovalbumin gene and identification of potential ovalbumin mRNA precursors. Cell 15:671-685.

154. Ross, D. A., Yen, R.-W., and Chae, C.-B. (1982). Association of globin ribonucleic acid and its precursors with the chicken erythroblast nuclear matrix. Biochemistry 21:764-771.

155. Ross, J. (1976). A precursor of globin messenger RNA. J. Mol. Biol. 106:403-420.

156. Ross, J., and Knecht, D. A. (1978). Precursors of alpha and beta globin messenger RNAs. J. Mol. Biol. 119:1-20.

157. Rubin, G. M., Kidwell, M., and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. Cell 29:987-994.

158. Ryffel, G. U., Wyler, T., Muellener, D. B., and Weber, R. (1980). Identification, organization, and processing intermediates of the putative precursors of Xenopus vitellogenin messenger RNA. Cell 19:53-61.

159. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., and Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162:729-773.

160. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, B. R., Reyes, A. A., and Bonner, J. (1979). The rat serum albumin gene: Analysis of cloned sequences. Proc. Natl. Acad. Sci. 76:3256-3260.

161. Sargent, T. D., Yang, M., and Bonner, J. (1981). Nucleotide sequence of cloned rat serum albumin messenger RNA. Proc. Natl. Acad. Sci. 78:243-246.

162. Sheiness, D., and Darnell, J. E. (1973). Polyadenylic acid segment in mRNA becomes shorter with age. Nature 241:265-268.

163. Schibler, U., Marcu, K. B., and Perry, R. P. (1978). The synthesis and processing of the messenger RNAs specifying heavy and light chain immunoglobulins in MPC-11 cells. Cell 15:1495-1509.

164. Schmid, C. W., and Jelinek, W. R. (1982). The Alu family of dispersed repetitive sequences. Science 216:1065-1070.

165. Scott, R. W., and Tilghman, S. M. (1983). Transient expression of a mouse α-fetoprotein minigene: Deletion analysis of promoter function. Mol. Cell. Biol. 3:1295-1309.

166. Sell, S., Becker, F. F., Leffert, H. L., and Watabe, H. (1976). Expression of an oncodevelopmental gene product (α-fetoprotein) during fetal development and adult oncogeneis. Cancer Res. 36:4239-4249.

167. Sell, S., and Gord, D. (1973). Rat α-fetoprotein--II. Refinement of radioimmunoassay for detection of 1 ng rat α_1F. Immunochem. 10:439-442.

168. Sell, S., Thomas, K., Michalson, M., Sala-Trepat, J., and Bonner, J. (1979). Control of albumin and α-fetoprotein expression in rat liver and in some transplantable hepatocellular carcinomas. Biochim. Biophys. Acta 564:173-178.

169. Sharp, P. A. (1981). Speculations on RNA splicing. Cell 23: 643-646.

170. Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. Cell 9: 645-653.

171. Shen, C.-K. J., and Maniatis, T. (1980). The organization of repetitive sequences in a cluster of rabbit β-like globin genes. Cell 19: 379-391.

172. Singer, M. F. (1982). SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. Cell 28: 433-434.

173. Sittman, D. B., Graves, R. A., and Marzluff, W. F. (1983). Structure of a cluster of mouse histone genes. Nucl. Acids Res. 11:6679-6697.

174. Small, D., Nelkin, B., and Vogelstein, B. (1982). Non-random distribution of repeated DNA sequences with respect to supercoiled loops and the nuclear matrix. Proc. Natl. Acad. Sci. 79:5911-5915.

175. Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

176. Strair, R. K., Skoultchi, A. I., and Shafritz, D. A. (1977). A characterization of globin mRNA sequences in the nucleus of duck immature red blood cells. Cell 12:133-141.

177. Strair, R. K., Yap, S. H., and Shafritz, D. A. (1977). Use of molecular hybridization to purify and analyze albumin messenger RNA from rat liver. Proc. Natl. Acad. Sci. 74:4346-4350.

178. Strauss, A. W., Donohoe, A. M., Bennett, C. D., Rodkey, J. A., and Alberts, A. W. (1977). Rat liver preproalbumin: *In vitro* synthesis and partial amino acid sequence. Proc. Natl. Acad. Sci. 74:1358-1362.

179. Sutcliffe, J. G. (1978). pBR322 restriction map derived from the DNA sequence: Accurate DNA size markers up to 4361 nucleotide pairs long. Cold Spring Harbor Symp. Quant. Biol. 43: 77-90.

180. Taylor, J. M., and Schimke, R. T. (1974). Specific binding of albumin antibody to rat liver polysomes. J. Biol. Chem. 249: 3597-3601.

181. Thomas, C. A. (1971). The genetic organization of chromosomes. Ann. Rev. Genet. 5:237-256.

182. Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. 77:5201-5205.

183. Tilghman, S. M., and Belayew, A. (1982). Transcriptional control of the murine albumin/α-fetoprotein locus during development. Proc. Natl. Acad. Sci. 79:5254-5257.

184. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Natl. Acad. Sci. 75:1485-1489.

185. Tsai, M.-J., Ting, A. C., Nordstrom, J. L., Zimmer, W., and O'Malley, B. W. (1980). Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. Cell 22:219-230.

186. van Eekelen, C. A. G., and van Venrooij, W. J. (1981). hnRNA and its attachment to a nuclear protein matrix. J. Cell. Biol. 88:554-563.

187. Vogelstein, B., Pardoll, D. M., and Coffey, D. S. (1980). Supercoiled loops and eukaryotic DNA replications. Cell 22: 79-85.

188. Wahl, G. M., Stern, M., and Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. 76:3683-3687.

189. Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299:691-698.

190. Watanabe, A., Miyazoki, M., and Taketa, K. (1976). Differential mechanisms of increased α-fetoprotein production in rats following carbon tetrachloride injury and partial hepatectomy. Cancer Res. 36:2171-2175.

191. Watanabe, A., Miyazaki, M., and Taketa, K. (1976). Prompt elevation of rat serum α-fetoprotein by acute liver injury following a single injection of ethionine. Int. J. Cancer 17:518-524.

192. Weinberg, R. A. (1973). Nuclear RNA metabolism. Ann. Rev. Biochem. 42:329-354.

193. Weiringa, B., Meyer, F., Reiser, J., and Weissmann, C. (1983). Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature 301:38-43.

194. Winberg, G., and Hammarskjold, M.-L. (1980). Isolation of DNA from agarose gels using DEAE-paper. Application to restriction site mapping of adenovirus type 16 DNA. Nucl. Acids Res. 8: 253-264.

195. Yang, V. W., Lerner, M. R., Steitz, J. A., and Flint, S. J. (1981). A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. Proc. Natl. Acad. Sci. 78:1371-1375.

196. Young, P. R., Scott, R. W., Hamer, D. H., and Tilghman, S. (1982). Construction and expression in vivo of an internally deleted mouse α-fetoprotein gene: Presence of a transcribed Alu-like repeat within the first intervening sequence. Nucl. Acids Res. 10:3099-3116.

197. Young, R. A., Hagenbuchle, O., and Schibler, U. (1981). A single mouse α-amylase gene specifies two different tissue-specific mRNAs. Cell 23:451-458.

198. Zaug, A. J., Grabowski, P. J., and Cech, T. R. (1983). Autocatalytic cyclization of an excised intervening sequence RNA is cleavage-ligation reaction. Nature 301:578-583.

199. Zeevi, M., Nevins, J. R., and Darnell, J. E. (1981). Nuclear RNA is spliced in the absence of poly(A) addition. Cell 26: 39-46.

200. Ziff, E. B. (1980). Transcription and RNA processing by the DNA tumor viruses. Nature 287:491-499.

201. Ziff, E. B., and Evans, R. M. (1978). Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit. Cell 15:1463-1475.

VITA

Harry Ratrie III was born in Baltimore, Maryland on July 27, 1945. He attended elementary and secondary schools in Towson, Maryland and was graduated from Towson Senior High School in June 1963. The following September he entered the Virginia Military Institute in Lexington, Virginia. In June 1967 he received a Bachelor of Arts degree in Biology and accepted a commission as a Second Lieutenant in the United States Army. Following a two-year tour of duty with the army, he was employed by Microbiological Associates, Inc. for a period of five years. In September of 1975 he began graduate studies at The University of Tennessee, Oak Ridge Graduate School of Biomedical Sciences, from which he received the Doctor of Philosophy degree in March of 1984.