Paul, M., Goldsmith, M. R., Hunsley, J. R., & Kafatos, F. C. (1972) *J. Cell Biol. 55*, 653–680.
Perry, R. P., & Scherrer, K. (1975) *FEBS Lett. 57*, 73.
Revel, M., & Groner, Y. (1978) *Annu. Rev. Biochem. 47*, 1079–1126.
Shatkin, A. J. (1976) *Cell (Cambridge, Mass.) 9*, 645–653.
Stuart, S. E., & Rottman, F. M. (1973) *Biochem. Biophys. Res. Commun. 55*, 1001–1008.

5′ Domain and Nucleotide Sequence of an Adult Chicken Chromosomal β-Globin Gene†

Lawrence E. Day,* Abigail J. Hirst, Eugene C. Lai, Myles Mace, Jr., and Savio L. C. Woo*

ABSTRACT: A 6.1-kilobase chromosomal chicken DNA fragment containing an adult β-globin gene has been cloned from total *Eco*RI-digested chicken DNA. The complete β-globin gene is approximately 1600 nucleotides in length and is flanked by 0.9 and 3.6 kilobases of DNA in this cloned fragment at its 5′ and 3′ termini, respectively. The chicken β-globin contains two intervening sequences at the same locations as the mammalian globin genes. The CAP site of the globin gene was determined by electron microscopic and S1 nuclease mapping, and the nucleotide sequence at the 5′ end of the gene was established. The gene contains 80 nucleotides at its 5′-untranslated region. At 29 nucleotides preceding the CAP site is the Hogness box sequence GATAAAA. The nucleotide sequences surrounding this AT-rich region are extremely GC rich (70%). The occurrence of CG dinucleotides is unusually frequent in this region, resulting in the appearance, immediately preceding the Hogness box sequence, of six *Hpa*II sites (recognition sequence CCGG) which could be targets for DNA methylation in the genome. A stretch of 16 dG residues which is not present in other globin genes is located 194 nucleotides preceding the CAP site. However, limited nucleotide conservation at the region between the Hogness box and the CAP site was observed between the avian and mammalian β-globin genes, but not between the chicken β-globin gene and the chicken ovalbumin, ovomucoid, and conalbumin genes. Although there are additional sequence homologies between the avian and mammalian β-globin genes further 5′ from the Hogness boxes, none is as striking as the pentanucleotide sequence CCAAT located exactly at 45 nucleotides preceding the Hogness box sequence in the avian and mammalian β-globin genes but not the other chicken genes.

Recombinant DNA technology has permitted the dissection of the eucaryotic genome. Structural and sequencing analyses after gene isolation by molecular cloning have established that mammalian globin genes are comprised of three mRNA coding segments separated by two intervening sequences (Tilghman et al., 1978; Tiemeier et al., 1978; Leder et al., 1978; Konkel et al., 1979; Nishioka & Leder, 1979; Lawn et al., 1978; Lacy et al., 1979; Hardison et al., 1979; van den Berg et al., 1978; Bernards et al., 1979). More recently, such a structure has also been observed in an adult chicken β-globin gene (Dodgson et al., 1979; Ginder et al., 1979). In the past, we and others have cloned the chromosomal ovalbumin and ovomucoid genes from total chicken DNA and shown that both of these estrogen-inducible genes in the chick oviduct contain seven intervening sequences (Woo et al., 1978; Dugaiczyk et al., 1978, 1979; Mandel et al., 1978; Catterall et al., 1979; Lai et al., 1979). Nucleotide sequencing analysis at the 5′ ends of these two genes has revealed some similarities (Lai et al., 1979). In order to compare the nucleotide sequences flanking the 5′ ends of the hormone-inducible genes with that of a steroid irresponsive chicken gene and those flanking the 5′ ends of mammalian and avian globin genes, we have cloned an adult chicken β-globin gene and determined its 5′-nucleotide sequence.

Materials and Methods

Materials. White Leghorn laying hens were purchased from Rich-Glo Farm, La Grange, TX, and the animals were made anemic by daily injection of phenylhydrazine. On the seventh day, the animals were sacrificed, and reticulocytes were obtained from plasma by low-speed centrifugation. Their livers were rinsed with cold 0.9% NaCl solution and frozen in liquid nitrogen immediately after dissection. Liquified phenol and reagent grade chemicals were purchased from Fisher Scientific Co. Formamide was obtained from Fluka. Restriction enzymes, T4 DNA ligase, and sequence grade polyacrylamide were purchased from Bethesda Research Laboratories. S1 nuclease was from Miles Laboratories. DNase and bacterial alkaline phosphatase were from Worthington. *Escherichia coli* DNA polymerase I and T4 polynucleotide kinase were from Boehringer Mannheim. Radioactive deoxyribonucleoside triphosphates were purchased from Amersham.

Methods. Reverse-Phase Column Chromatography. The method has been successfully used to enrich the mouse β-globin gene from total mouse DNA for its subsequent cloning (Tilghman et al., 1977). One hundred milligrams of *Eco*RI-digested chicken liver DNA in 10 mM Tris[1]–acetate, pH 7.4, 1.5 M NaOAc, and 0.1 mM EDTA was chromatographed on a RPC-5 column (1 × 50 cm) by the procedure of Hardies & Wells (1976). The DNA was eluted from the column in

† From the Howard Hughes Medical Institute and Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030. *Received April 22, 1980; revised manuscript received November 6, 1980.* This work was partly supported by the Baylor Center for Population Research. S.L.C.W. is an investigator of the Howard Hughes Medical Institute.

[1] Abbreviations used: Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; $NaDodSO_4$, sodium dodecyl sulfate; kb, kilobase; bp, base pairs; cDNA, complementary deoxyribonucleic acid; rRNA, ribosomal ribonucleic acid.

0006-2960/81/0420-2091$01.25/0

10-mL fractions by using 1 L of a linear gradient containing 1.5–1.8 M NaOAc in the same buffer. DNA from each fraction was precipitated with alcohol and redissolved in 0.5 mL of 10 mM Tris-HCl, pH 7.6, and 1 mM EDTA.

Assay of β-Globin DNA-Containing Sequences by Southern Hybridization. Two micrograms of DNA from alternate fractions from the RPC-5 column was electrophoresed on a 1% agarose gel in Tris–acetate and transferred onto nitrocellulose filters by the method of Southern (1975), as described previously by Lai et al. (1978). The filters were hybridized with a nick-translated [^{32}P]DNA probe of pHb1001, which is a recombinant plasmid constructed with pMB9 DNA and a double-stranded DNA synthesized from adult chicken β-globin mRNA. The plasmid was kindly provided by Winston Salser (UCLA). The filters were subjected to radioautography after extensive washing, as described by Lai et al. (1978).

Labeling of DNA Probes by Nick Translation. The procedure used was based on that of Maniatis et al. (1975) and Rigby et al. (1977). The reaction was performed in a final volume of 25 μL containing 1 μg of DNA in 50 mM Tris-HCl (pH 7.8), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 5 μg of BSA, 30 μM dATP and dGTP, 5 μM [^{32}P]dCTP, and [^{32}P]TTP. The mixture was prepared on ice, and 0.1 ng of DNase was added. After being incubated for 1 min at room temperature, the mixture was immediately cooled in an ice water bath. Ten units of *E. coli* DNA polymerase I was then added, and the reaction mixture was incubated at 14 °C. The reaction was stopped after 90 min of incubation by the addition of 5 μL of 0.1 M Na_2EDTA and 5 μL of 10% $NaDodSO_4$. Unincorporated dNTPs were separated from the labeled DNA by gel filtration. Radioactive fractions eluting in the void volume were pooled and precipitated by 2 volumes of ethanol. The precipitate was resuspended in 0.5 mL of 10 mM Tris-HCl (pH 7.6) and 1 mM Na_2EDTA. It was made to contain 0.1 N NaOH and incubated at room temperature for 2 h. The ^{32}P-labeled DNA was chilled, neutralized with HCl, and kept at −20 °C. The specific activity of the DNA probes was generally (1–2) × 10^8 cpm/μg.

Preparative Agarose Gel Electrophoresis. Preparative agarose gel electrophoresis was performed as described previously (Lai et al., 1979). Briefly, 200 μg of restriction endonuclease digested DNA was applied to a 0.6-cm thick agarose slab gel by using an electrophoresis apparatus by E-C Apparatus Corp. After electrophoresis, the agarose gel strip containing the desired DNA fragment was excised and crushed by passing it through a 21-gauge needle. The crushed gel was suspended in 2 volumes of extraction buffer (50 mM Tris-HCl, pH 7.6, 100 mM NaCl, and 1 mM Na_2EDTA) and allowed to stand at room temperature for 2 days. Agarose was pelleted by centrifuging the gel suspension in an SW40 rotor for 3 h at 32 000 rpm. The supernatant was extracted on ice with an equal volume of redistilled phenol saturated with extraction buffer. After centrifugation, the aqueous layer was precipitated with ethanol. The precipitate was resuspended in a small volume of extraction buffer and further extracted with phenol on ice until no interphase was present. The DNA was recovered after precipitation with ethanol, and the yield ranged between 50% and 80% of the starting material.

Cloning of an Adult Chicken β-Globin Gene. Escherichia coli LE392/ThyA and the bacteriophage λgtWES·λB were kindly provided by Dr. Philip Leder (National Institutes of Health). The bacteriophage was digested with *Eco*RI and the middle λB fragment was removed from λgtWES DNA by preparative agarose gel electrophoresis. The subsequent cloning experiments were performed in a P3 facility in accordance with the NIH guidelines for recombinant DNA research.

Two micrograms of the λgtWES DNA and 1 μg of the partially purified globin DNA were incubated with 1 unit of T4 DNA ligase in a total reaction volume of 20 μL at 12 °C for 15 h. Ten microliters of the ligation mixture was used for cloning by in vitro packaging according to the procedure of F. Blattner as reported previously by Dugaiczyk et al. (1979). Briefly, lysogens were induced and cell extracts were prepared by alternate freezing and thawing and by sonication. The two protein extracts were mixed with partially purified protein A and incubated at 37 °C for 60 min with the ligated DNA in a final volume of 50 μL. The packaging mixture was plated in a 5-μL aliquots in soft agar on ten agar plates. Approximately, 12 000 phage plaques were generated in this experiment.

The phage plaques were screened for the recombinant containing the β-globin gone by hybridization with the pHb1001 probe, utilizing an amplification procedure previously described by Woo (1979).

R-Loop Analysis by Electron Microscopy. Hybrids of the cloned β-globin gene with chicken globin mRNA were formed with 10 μg/mL DNA and 20 μg/mL RNA in 70% deionized formamide containing 100 mM Tris-HCl, pH 7.6, 10 mM Na_2EDTA, and 200 mM NaCl. The mixture was heated at 80 °C for 5 min and incubated at 55 °C for 3 h. The reaction mixture was then permitted to cool slowly to room temperature (~2 h). Following incubation, samples were prepared for electron microscopy by diluting 0.1–0.5 μg of nucleic acids with 100 μL of a solution containing 70% formamide, 0.1 M Tris-HCl, pH 8.4, 0.01 M Na_2EDTA, and 100 μg/mL cytochrome *c*. The mixture was spread onto a hypophase of distilled water. Samples were picked up on collodion-coated 300 mesh copper grids, stained with uranyl acetate, rotary shadowed with platinum–palladium, and examined at 80 kV by using a Joel 100-C electron microscope. Molecules were measured by using a Neumonics electronic planimeter with φX174 RF DNA as the internal standard.

5′ End Labeling of DNA. pCβG6.1 DNA was digested to completion with *Hin*dIII, and the 2.03-kb fragment was purified by preparative agarose gel electrophoresis. The 5′-terminal phosphates on the DNA were removed by incubation at 68 °C with 1 unit of bacterial alkaline phosphatase in a reaction volume of 50 μL containing 0.01 M Tris-HCl, pH 7.8, 0.01 M $MgCl_2$, and 0.001 M dithiothreitol. After 30 min, the mixture was extracted with an equivalent volume of phenol presaturated with 0.05 M Tris-HCl, pH 7.4, and 0.001 M EDTA and precipitated with ethanol. The 5′-OH groups were subsequently labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. The pellet was redissolved in 20 μL of 50 μM glycine–NaOH, pH 9.5, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine, and 25% glycerol. The solution was added to 200 μCi of dried [γ-^{32}P]ATP, and 4 μL of T4 polynucleotide kinase (4000 units/mL) was added. The mixture was incubated at 37 °C for 30 min, and the reaction was terminated by phenol extraction.

Both 5′ termini of the DNA were labeled in this procedure. The DNA was subjected to digestion with *Eco*RI after removal of excess [^{32}P]ATP by gel filtration to obtain a probe with only one labeled terminus. The end-labeled DNA fragments were subsequently separated by preparative agarose gel electrophoresis.

S1 Nuclease Mapping. Hybridization between 50 ng of end-labeled probe (35 000 cpm) and 150 μg of poly(A)-containing chicken reticulocyte RNA was performed in a final

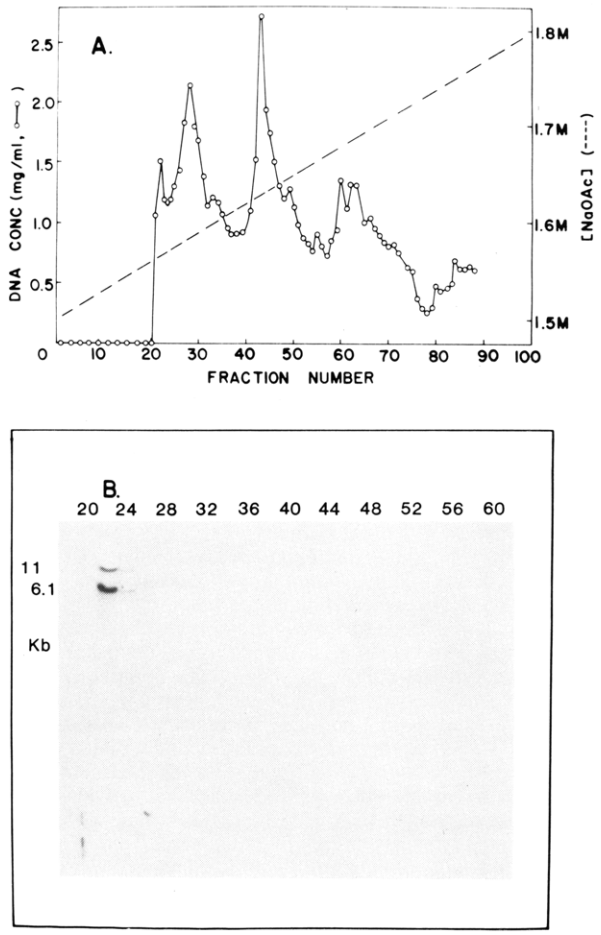

FIGURE 1: RPC-5 column chromatography of total *Eco*RI-digested chicken DNA. (Panel A) Elution profile of 100 mg of chicken DNA on an 1 × 50 cm column. (Panel B) Radioautogram of a Southern agarose gel blot containing 2 μg of DNA from various fractions of the RPC-5 column. The hybridization probe was ^{32}P-labeled pHb1001 DNA, a recombinant plasmid containing a cDNA insert of an adult chicken β-globin gene. The two hybridization bands in fraction 23 were 11 and 6.1-kb in length, as indicated on the left-hand side of the radioautogram.

volume of 100 μL that contained 70% deionized formamide, 0.3 M NaCl, 0.01 M Tris-HCl, pH 7.5, and 0.001 M EDTA. DNA was denatured at 100 °C for 3 min prior to incubation at 54 °C for 18 h. Nucleic acid in the reaction mixture was precipitated with ethanol for subsequent S1 nuclease digestion in a final volume of 500 μL that contained 0.2 M sodium acetate, pH 4.5, 0.4 M NaCl, 0.002 M $ZnCl_2$, 12.5 μg of denatured calf thymus DNA, and 3400 units of S1 nuclease. The reaction mixture was incubated at 37 °C for 30 min and was terminated by ethanol precipitation. The pellet was resuspended in 10 μL of 0.1 N NaOH. After incubation at 68 °C was continued for 5 min, the alkali-treated sample was applied to a 4% polyacrylamide gel containing 7 M urea and electrophoresed at 200 V for 2–3 h. The gel was then subjected to radioautography.

DNA Sequencing. Chemical degradation of end-labeled DNA was performed according to the method of Maxam & Gilbert (1977), and polyacrylamide gel electrophoresis using the thin gel was performed according to Sanger & Coulson (1978).

Results

Cloning of a Chicken Chromosomal Adult β-Globin Gene. pHb1001 is a recombinant plasmid constructed by inserting a cDNA copy of an adult chicken β-globin mRNA into the

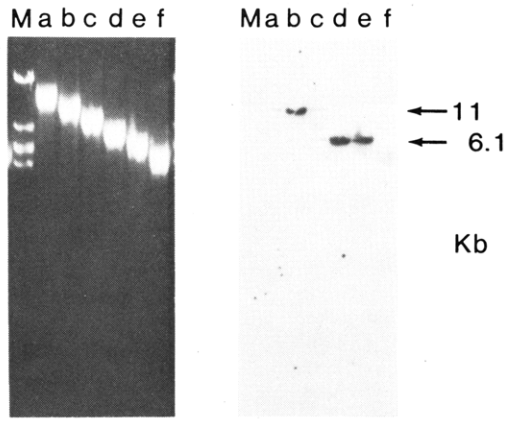

FIGURE 2: Enrichment of the globin gene fragments by preparative agarose gel electrophoresis. (Left panel) Ethidium bromide stain of an agarose gel containing various fractions (lanes a–f) recovered from a preparative agarose gel. Lane M contains *Eco*RI-digested λ DNA. (Right panel) Radioautogram of the same agarose gel after hybridization with ^{32}P-labeled pHb1001 DNA.

*Eco*RI site of plasmid pMB9 DNA. When this recombinant plasmid was used as a probe, two hybridization bands at 11 and 6.1 kb were detected in *Eco*RI-digested total chicken DNA. These two hybridizable fragments were enriched from total DNA by RPC-5 column chromatography, and the elution profile of the column is shown in Figure 1A. Southern hybridization analysis demonstrated that both β-globin gene containing fragments eluted at the earliest fractions from the column (Figure 1B), and an enrichment of about 20-fold was achieved. The fragments containing β-globin genes were further enriched by preparative agarose gel electrophoresis. DNA from fraction 23 of the RPC-5 column was electrophoresed on a 1% agarose gel which was subsequently cut into 5-mm slices. DNA recovered from each gel slice was analyzed for globin gene content. The ethidium staining pattern of DNA from individual gel slices on an analytical agarose gel is shown in Figure 2 (left panel). Southern hybridization analysis shows that the 6.1-kb globin gene fragment is particularly enriched in fractions d and e from the preparative gel (Figure 2, right panel). DNA from these gel fractions were then allowed to ligate with the phage DNA arms of λgtWES vector and cloned in *E. coli* LE392 by in vitro packaging. Two signals were obtained from 12000 recombinant phages. *Eco*RI digestion of these phage DNAs generated λgtWES DNA and an additional 6.1-kb fragment which hybridized with ^{32}P-labeled pHb1001 DNA (not shown). The 6.1-kb chicken DNA isolated from these recombinant phages was then recloned in the *Eco*RI site of the plasmid vector pBR322, and the recombinant was designated pCβG6.1.

Molecular Structure of the Chicken β-Globin Gene. The molecular structure of the cloned β-globin gene was analyzed by electron microscopy. *Eco*RI-digested pCβG6.1 DNA was allowed to hybridize with poly(A)-containing chicken reticulocyte RNA under R-loop conditions. The electron micrograph of a hybrid molecule and the corresponding line drawing are shown in Figure 3. The molecule indicates the presence of at least two structural gene regions separated by an intervening sequence (IVS B). The small intervening sequence (IVS A) as reported by Ginder et al. (1979) did not form a stable duplex under the hybridization conditions employed and is present as a single-stranded DNA loop that has the appearance of a thick region in the larger R loop within these molecules. The poly(A) tail of the mRNA is clearly present and is located adjacent to the small R loop. The duplexed and looped regions of 20 individual molecules were determined by using a numonic

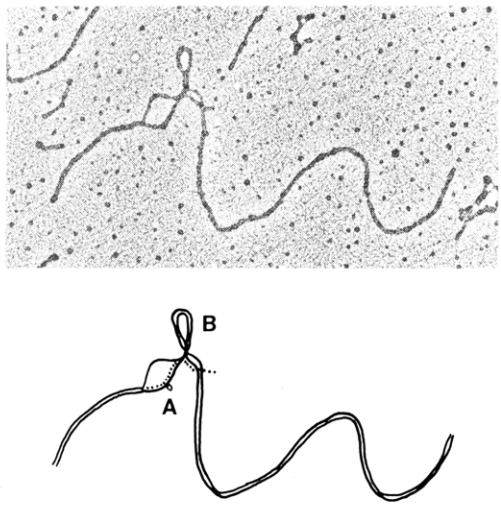

FIGURE 3: Electron micrograph and the corresponding line drawing of a hybrid molecule formed between adult chicken β-globin mRNA and the cloned 6.1-kb chicken fragment containing a chromosomal β-globin gene. Hybridization was carried out under R-loop conditions, and electron microscopy was performed as described under Materials and Methods. The poly(A) tail on the mRNA and the large intervening sequence B are clearly visible. The small intervening sequence A was too small to form a stable DNA duplex. It is present as a single-stranded DNA loop within the large R-loop and has the appearance of a thickened area within the RNA/DNA duplex. This thickened area is present at exactly the same location in ten individual hybrid molecules.

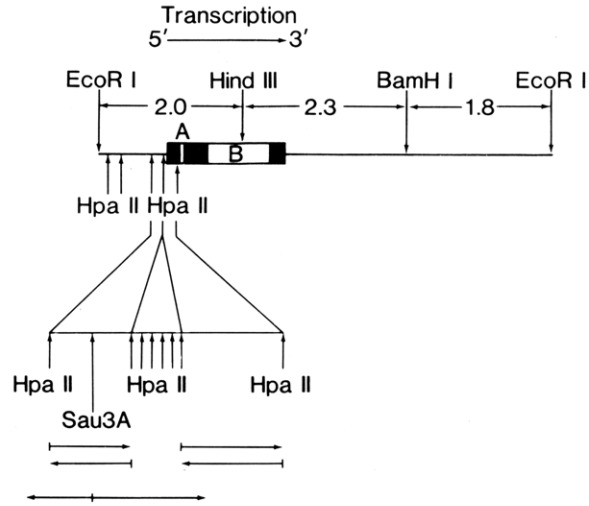

FIGURE 4: Physical map of the cloned 6.1-kb chicken DNA fragment. The DNA was cleaved by *Hin*dIII into two fragments of 4.1 and 2.0 kb in length, and the former fragment was cleaved by *Bam*HI into two fragments of 2.3 and 1.8 kb in length (not shown). The unique *Hin*dIII site in this fragment must be internal to the β-globin gene since both the 2.0-kb *Eco*RI/*Hin*dIII and the 2.3-kb *Hin*dIII/*Bam*HI fragments hybridized to pHB 1001 DNA. The transcription orientation of the gene was determined by the fact that the poly(A) tail on the mRNA is flanked by the longer DNA arm in the cloned 6.1-kb DNA (Figure 3). The approximate positions of the three structural gene regions and the two intervening sequences on the fragment were determined from electron microscopic analysis of ten individual hybrid molecules with β-globin mRNA. The lengths of the flanking sequences and various gene segments on each hybrid molecule were determined by using a numonic map measure. Ten independent determinations were averaged and employed to construct the β-globin gene structure. The size of intervening sequence A, however, cannot be accurately determined on these molecules and is only an approximation on the schematic drawing. The five *Hpa*II sites within the 2.0-kb *Eco*RI/*Hin*dIII fragment were determined by partial *Hpa*II digestion of the fragment labeled only at the *Hin*dIII end followed by polyacrylamide gel electrophoresis and radioautography (data not shown). The detection of six bands at 2000, 1860, 1665, 1165, 1050, and 900 bp in length indicated the presence of five *Hpa*II sites within the DNA fragment as shown. Indeed, the 2.0-kb *Eco*RI/*Hin*dIII fragment was cleaved by *Hpa*II into six fragments of 900, 500, 195, 150, 140, and 105 bp in length (Figure 5). Upon DNA sequence analysis, however, the *Hpa*II site separating the 105- and 150-bp *Hpa*II fragments was found to consist of six *Hpa*II sites. Strategy used to sequence the 5′ flanking sequence of the globin gene is shown in the lower portion of the diagram.

map measure, and the average was used to construct a β-globin gene map (Figure 4). The orientation of the globin gene in the cloned chicken DNA was from left to right as shown in the map because of the location of the poly(A) tail of the mRNA in the electron micrograph. The large loop containing the first two structural gene segments plus the first intervening sequence is approximately 600 base pairs (bp) in length and is separated from the 200-bp third structural gene segment by a 900-bp intervening sequence. The entire β-globin gene is flanked in the cloned 6.1-kb chicken DNA fragment by approximately 900 and 3600 bp of DNA at its 5′ and 3′ termini, respectively.

5′ Domain of the Chicken β-Globin Gene. The positions of the small intervening sequence (IVS A) and the CAP site of the gene were determined by S1 nuclease mapping. The 2.03-kb *Hin*dIII fragment from pCβG6.1 DNA was 5′ end labeled with ^{32}P by using successively alkaline phosphatase and polynucleotide kinase in the presence of [γ-^{32}P]ATP. The labeled DNA fragment was cleaved with *Eco*RI, and the 2.0-kb *Hin*dIII/*Eco*RI fragment containing the 5′ half of the globin gene was isolated by preparative gel electrophoresis. This DNA fragment, labeled only at the *Hin*dIII site present in IVS B, was then denatured thermally and allowed to hybridize with total poly(A)-containing RNA isolated from chicken reticulocytes. The reaction mixture was subsequently treated with S1 nuclease, and the residual-labeled fragments were detected by polyacrylamide gel electrophoresis followed by radioautography. There should be a minimum of four molecular species of globin RNA in the preparation (Figure 5): (1) precursor RNA with both intervening sequences; (2) precursor RNA with IVS A removed; (3) precursor RNA with IVS B removed; and (4) mature mRNA with both intervening sequences removed. Although the latter two species would have hybridized with the end-labeled DNA probe, the labeled end would not be protected from S1 nuclease due to the absence of IVS B in the RNA molecules (Figure 5). When the end-labeled DNA probe was hybridized to the precursor RNA

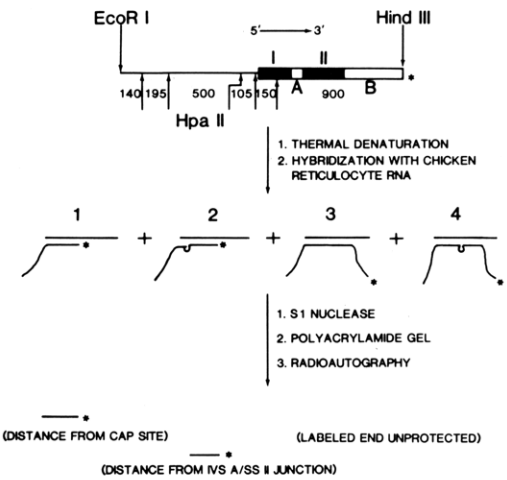

FIGURE 5: Schematic representation of the strategy to determine the 5′ domain of the β-globin gene by S1 nuclease mapping. The labeled *Hin*dIII end is indicated by the asterisk.

without IVS A, the looped DNA region corresponding to IVS A would be cleaved by S1 nuclease, and the size of the re-

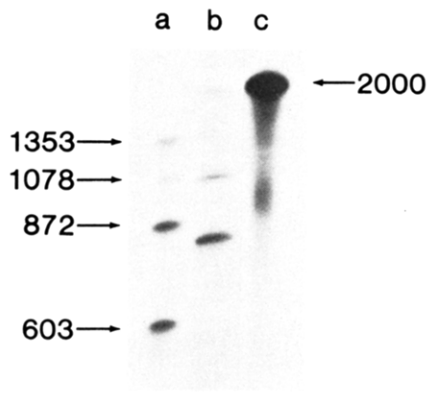

FIGURE 6: Radioautogram of a polyacrylamide gel in S1 nuclease mapping experiment described in Figure 5. (Lane a) *Hae*III-digested φX174 DNA after end labeling. Sizes of the individual DNA fragments in nucleotides are shown on the left-hand side of the radioautogram. (Lane b) End-labeled 2.0-kb *Eco*RI/*Hin*dIII fragment after hybridization with chicken reticulocyte RNA and S1 nuclease digestion. (Lane c) End-labeled fragment without S1 nuclease digestion.

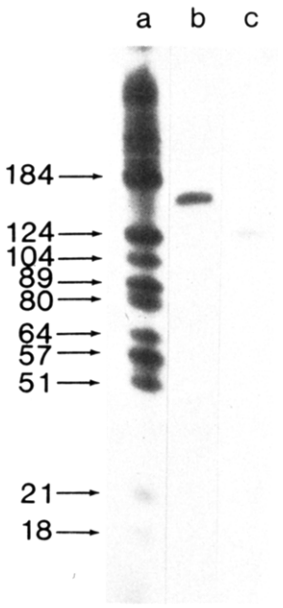

FIGURE 7: Radioautogram of polyacrylamide gel in a second S1 nuclease mapping experiment. Individual *Hpa*II fragments from the 2.0-kb *Eco*RI/*Hin*dIII fragment were end labeled and hybridized with chicken reticulocyte RNA. (Lane a) *Hae*III-digested pBR322 DNA after end labeling. (Lane b) 150-bp *Hpa*II fragment after end labeling. (Lane c) Labeled *Hpa*II fragment after hybridization with chicken reticulocyte RNA and S1 nuclease digestion.

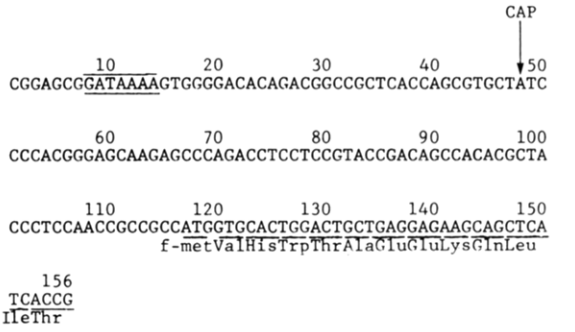

FIGURE 8: Complete nucleotide sequence of the *Hpa*II DNA fragment containing the 5′ end of the β-globin gene. The entire fragment is 156 nucleotides in length. The 3′-terminal 36 nucleotides corresponded exactly to the first 12 amino acids in adult chicken β-globin. Immediately preceding the codon sequence is the initiation codon ATG. The CAP site of the gene was determined only by S1 nuclease mapping.

sulting DNA fragment should reveal the distance of IVS A from the *Hin*dIII site. Indeed, a band of 830 nucleotides in length is evident in the radioautogram (Figure 6, lane b).

Since the precursor RNA containing both intervening sequences in collinear with the DNA probe, the entire 5′ globin gene portion will be protected from S1 nuclease, and the size of the resulting labeled DNA band should reveal the distance between the 5′-capping site of the gene and the unique internal *Hin*dIII site. An additional band at about 1050 nucleotides is evident in the radioautogram (Figure 6). The location of the capping site of the mRNA as determined by S1 nuclease mapping agrees well with restriction and electron micrographic mapping of the gene (Figure 4).

There appeared to be five *Hpa*II sites in the 2.0-kb *Hin*dIII/*Eco*RI fragment which yields a total of six *Hpa*II fragments of 900, 500, 195, 150, 140, and 105 base pairs in length (Figure 5). One of these *Hpa*II sites should correspond to the *Hpa*II site present at amino acid codon 12/13 within the first structural gene segment (Richards et al., 1979). S1 nuclease mapping was performed by using the individually end-labeled *Hpa*II fragments as the hybridization probe to establish the exact location of the capping site on the gene. Only the 900- and 150-bp DNA fragments were protected from S1 nuclease digestion. The 900-bp fragment was internal within the gene, and the neighboring 150-bp fragment was reduced in size to 120 nucleotides (Figure 7). This experiment would localize the CAP site of the gene being located at 1020 bp from the unique *Hin*dIII site, and the result agrees well with the previous S1 nuclease mapping and electron microscopic data.

5′-Nucleotide Sequence of the Gene. Since the above experiment has localized the position of the 5′ end of the gene to be within the 150-bp DNA fragment, the nucleotide sequence of this fragment was established by the method of Maxam & Gilbert (1977). The fragment was labeled at the 5′ end and recut with *Hae*III into two fragments of 120 and 30 bp in length. The nucleotide sequences of the two fragments were then determined. In order to sequence across the *Hae*III site in the fragments, the 5′-labeled DNA fragment was subjected to a strand separation gel, and both DNA strands were again sequenced. The entire fragment is 156 nucleotides in length, and its sequence is shown in Figure 8. The nucleotide sequence at the 3′ end of the fragment contains codons corresponding to the first 12 amino acids of adult chicken β-globin. These codons follow immediately the initiation codon ATG at nucleotides 117–119. Preceding the initiation codon is a noncoding region of the gene, which extends at least to nucleotide 68 since the sequence of this region is identical with that of a cDNA clone reported by Richards et al. (1979). The exact 5′ end of the gene, however, cannot be pinpointed yet on the chromosomal gene sequence since the entire sequence of the adult chicken β-globin mRNA has not been established. But our S1 nuclease mapping data would suggest that the 5′ end is positioned at around nucleotide 30 of the fragment. If the mRNA coding sequence does start with the dinucleotide AC at nucleotides 37–38, the 5′ untranslated region would be 80 nucleotides in length. This length is similar to the 76 nucleotides present in the chicken conalbumin gene but 27–28 nucleotides longer than those of mouse and rabbit adult β-globin genes. Similar to other eucaryotic genes, there appears to be a small palindromic structure at the 5′ end of the mRNA and two regions of complementarity with the 3′-terminal sequence of 18S mRNA (Figure 9).

For comparison of the 5′-flanking nucleotide sequence between the chicken β-globin gene with those of the mammalian

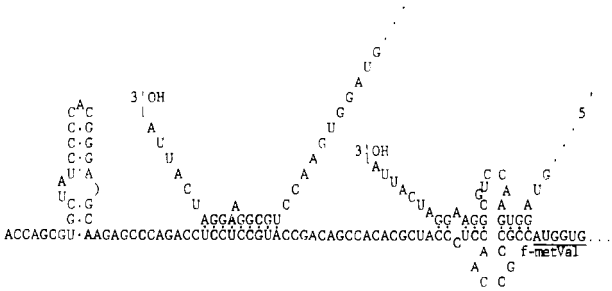

FIGURE 9: A possible structure of β-globin mRNA at the 5′ end. Immediately preceding the initiation codon AUG in the 5′ untranslated region exists some nucleotide complementarity with the 3′ end of the 18S rRNA. A second region of such complementarity is present 17 nucleotides further 5′ from the first region. These regions are then preceded by a small palindromic structure located close to the 5′ end of the molecule.

β-globin genes and the estrogen responsive genes in the chick oviduct, the sequence of the neighboring 105-bp *Hpa*II fragment was determined. This fragment contains a *Sau*3A site which was labeled, and the appropriate fragments were sequenced in order to establish the orientation of the 105-bp *Hpa*II fragment and to cross the *Hpa*II site separating the 105- and the 150-bp *Hpa*II fragments (Figure 4). Suprisingly, such an experiment revealed that there were actually a total of six *Hpa*II sites scattered over a 70-bp region of DNA (Figure 10), which is rather unusual since the *Hpa*II recognition sequence is CCGG and CG dinucleotides are uncommon in the eukaryotic genome.

The nucleotide sequence preceding the chicken β-globin gene is extremely GC rich (70%), and there is a stretch of 16 dG's located at 194 nucleotides preceding the CAP site of the gene which is not present in the other genes. At 29 nucleotides preceding the mRNA coding portion of the gene is the AT-rich heptanucleotide sequence GATAAAA (Figure 10). A similar sequence at the same locus is also present in all these other genes. Other than this AT-rich region, there is no other striking sequence homology between the chicken genes. However, in the region covering the first 12–13 nucleotides immediately preceding the gene, there are significant sequence similarities between the chicken and the mammalian β-globin genes. More interestingly, there is a pentanucleotide sequence CCAAT located exactly at 45 nucleotides proceding the Hogness boxes of the β-globin genes but not the other chicken genes at this position.

Discussion

By use of pHB1001 DNA as a specific hybridization probe, a 6.1-kb chicken DNA fragment containing an adult β-globin gene has been cloned in the λgtWES DNA vector after enrichment from *Eco*RI-digested total chicken DNA by a combination of RPC-5 column chromatography and preparative agarose gel electrophoresis. The entire β-globin gene is present within this cloned *Eco*RI fragment. The gene contains three structural segments separated by two intervening sequences. Although the exact locations of the two intervening sequences within the gene have not been established by nucleotide sequencing, electron microscopic analysis and S1 nuclease mapping experiments have indicated that their positions correspond to the intervening sequences present within the mammalian globin genes.

The adult chicken β-globin mRNA appears to contain a 5′ untranslated region that is 27–28 nucleotides longer than the mouse, rabbit, and human β-globin mRNAs (Konkel et al., 1979; Russel et al., 1976; Kafatos et al., 1977; Baralle, 1977; Proudfoot, 1977; Efstratiadis et al., 1977; van den Berg et al., 1978; Marotta et al., 1977). As in other eucaryotic genes, the first ATP triplet in the structural chicken β-globin sequence is used as the initiation codon for translation. In many other eucaryotic messenger RNAs, there exist segments within the 5′ untranslated region that are complementary to the 3′-terminal sequence of eucaryotic 18S rRNA which might be in-

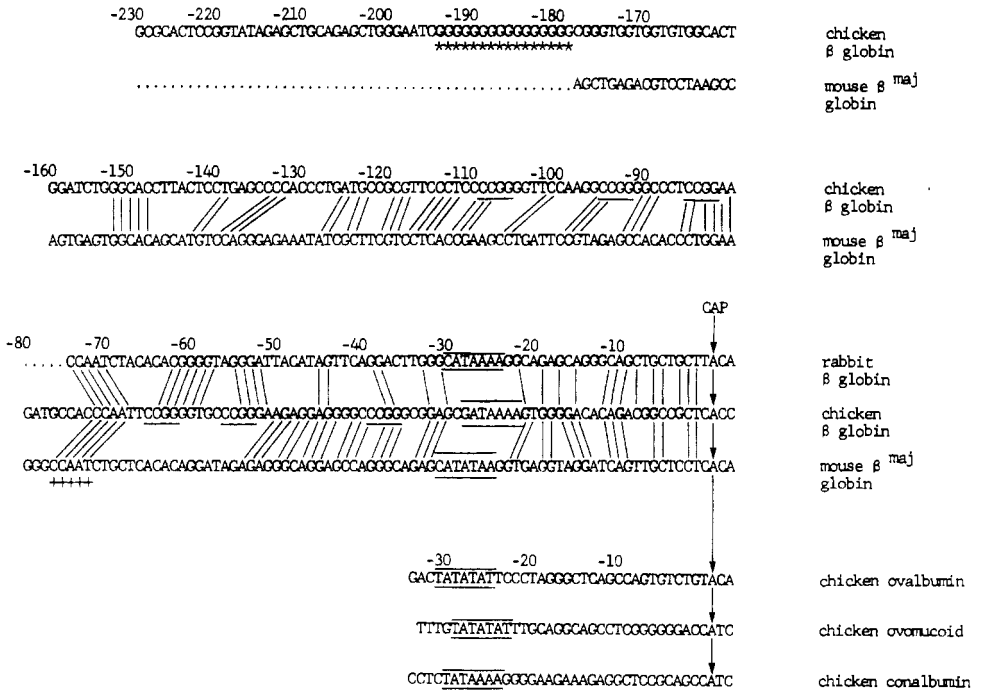

FIGURE 10: Comparison of the nucleotide sequences immediately flanking the adult chicken β-globin gene, the adult mammalian β-globin genes (Konkel et al., 1979; Hardison et al., 1979), and the hormone-responsiveness genes in the chicken oviduct (Lai et al., 1979; Cochet et al., 1979). All sequences are aligned at the CAP sites. The multiple recognition sequences for *Hpa*II in this region of the chicken β gene are underlined, and the poly(dG) region is indicated by the asterisks. The common Hogness box sequences in the genes are boxed, and conserved nucleotides in the globin genes are shown by straight lines between the gene sequences. The common pentanucleotide sequence CCAAT at 45 nucleotides preceding the Hogness box sequence is indicated by + signs.

volved in ribosome binding and translation initiation (Kozak, 1978; Hagenbiichle et al., 1978). Two such segments are apparent in the adult chicken β-globin (Figure 9). Only one such segment is present in both the chicken ovalbumin and ovomucoid genes with 5′ untranslated regions of 65 and 53 nucleotides in length, respectively. The chicken conalbumin gene with a comparable length of 76 nucleotides in the 5′ untranslated region also contains two such segments, although the complementarity is not as extensive as in the β-globin gene (Cochet et al., 1979). Furthermore, such complementary regions in the β-globin gene are preceded by a small palindromic structure which is also present in many other eucaryotic genes including those coding for chicken ovalbumin and conalbumin; rabbit mouse, and human α- and β-globins (Cochet et al., 1979; Efstratiadis et al., 1977; Baralle, 1977). Although the functional role(s) of such palindromic structures has not yet been established, such structures are present in control regions of many procaryotic operons and bacteriophages (Maniatis et al., 1975; Gilbert & Maxam, 1973; Bennett et al., 1976).

Another important feature of the chicken β-globin gene is the heptanucleotide sequence GATAAAA that precedes the mRNA coding region by 29 base pairs and is surrounded on both sides by GC-rich regions. A similar sequence has been found in many other eucaryotic genes at the same position, and a promoter function has been implicated, owing to sequence similarity to that of the Pribnow box in procaryotic genes. Other than this Hogness box structure, there is no apparent homology in the immediate 5′-flanking sequences between the four chicken genes. On the other hand, there is some conservation of nucleotide sequence in this region between the chicken and mammalian globin genes. The "capping base" sequence (CAGTTGCTCCTCAC) present in mouse β major and β minor, rabbit β-globin genes and a major adenovirus mRNA (Nishioka & Leder, 1979) appear to be also present in the chicken β-globin gene. The pentanucleotide TTGCT or CTGCT present in the mammalian α- and β-globin genes within the capping sequence is not rigidly conserved in the chicken β-globin gene, although the dinucleotide GC is still present at nucleotide position 7. Since the capping base sequence is not present in the mouse α-globin gene and other chicken genes, its functional role(s) in gene expression and regulation has not been defined at the present time.

Additional sequence homologies between the avian and the mammalian β-globin genes preceding the Hogness box are present. The homology of the 20 bp of DNA immediately 5′ from the Hogness box sequence appears to be more extensive between the chicken and the mouse genes and that between 20 and 30 bp of DNA preceding the box appears to be more extensive between the chicken and the rabbit genes. Since the homologous sequences at this region of the three genes do not overlap, their functional role(s) in the differentiation or expression of the genes is not apparent. The only homologous region present in all three globin genes is the pentanucleotide CCAAT at 45 bp preceding the Hogness box. The distance between this pentanucleotide and the Hogness box is curiously constant in all three genes. In addition, the CCAAT pentanucleotide sequence is not only present in the mouse β minor gene and the human adult β-globin gene, but their locations are also 45 bp from the Hogness box sequence (Konkel et al., 1979; Lacy & Maniatis, 1980). The CCAAT sequence is also present in the mouse and human α-globin genes, but their positions are not exactly 45 bp from the Hogness boxes (Nishioka & Leder, 1979; Proudfoot & Maniatis, 1980). Although it has been reported that a similar sequence appears to be present in other eukaryotic genes (Benoist et al., 1980), these sequences not only differ from the pentanucleotide sequence by at least two nucleotides, but their locations from the Hogness boxes are also variable. Whether the CCAAT sequence is involved in the expression of the adult β-globin gene only in differentiated cells, however, cannot be established at the present time. It will be interesting to see if the embryonic β-globin genes also have the pentanucleotide sequence at that position. Its functional role in transcription of the gene, if any, may be established possibly by a combination of site-directed mutagenesis and gene expression studies either in a RNA polymerase II dependent system or in a DNA-mediated gene-transfer system.

The DNA sequences surrounding the Hogness box are extremely GC rich (70%) and are curiously enriched in CG dinucleotides, resulting in the presence of six *Hpa*II sites over a range of 70 nucleotides. Recent studies have correlated a general pattern of DNA methylation and gene expression, suggesting that gene transcription is associated with undermethylation at specific sites within the genes (Waalmijk & Flavell, 1978; Bird et al., 1979; Mandel & Chambon, 1980). Furthermore, it has been reported that a *Hpa*II site at close vicinity to the 5′ end of the adult chicken β-globin gene is completely unmethylated in cells which are expressing or have expressed this gene but is methylated partially in other cells (McGhee & Ginder, 1979). The analysis employed in the methylation study was by comparing the *Hpa*II and *Msp*I cleavage patterns of genomic DNA from different chicken tissues. The extent of methylation at the six *Hpa*II sites immediately flanking the 5′ end of the chicken β-globin gene is difficult to determine, since any one of the six sites not methylated would have resulted in the cleavage by *Hpa*II at this region and would be scored as unmethylated. By the same analysis, complete unmethylation at this region of the gene in cells that had expressed or is expressing this gene would mean a range of complete to partial unmethylation in actuality. The multiple CG dinucleotides present at the strategic position of the chicken adult β-globin gene could be targets for methylation of the DNA in the genome, which might be involved in the depression and activation of β-globin gene expression during erythroid cell differentiation.

Acknowledgments

We thank Dr. Winston Salser, Department of Biology, UCLA, for kindly providing the recombinant plasmid pHb1001, Dr. Achilles Dugaiczyk, Department of Cell Biology, Baylor College of Medicine, for making available to us the λ in vitro packaging components, and Wanda Beattie for discussion on DNA sequencing.

References

Baralle, F. (1977) *Cell* (*Cambridge, Mass.*) *10*, 549–558.

Bennett, G. N., Schweingruber, M. E., Brown, K. D., Squires, C., & Yanofsky, C. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2351–2355.

Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142.

Bernards, R., Little, P. F. R., Annison, G., Williamson, R., & Flavel, R. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4827–4831.

Bird, A. P., Taggard, M. H., & Smith, B. A. (1979) *Cell* (*Cambridge, Mass.*) *17*, 889–901.

Catterall, J. F., Stein, J. P., Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Mace, M. L., Means, A. R., & O'Malley, B. W. (1979) *Nature* (*London*) *278*, 323–327.

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., & Chambon, P. (1979) *Nature (London) 282*, 567–574.
Dodgson, J. B., Strommer, J., & Engel, J. D. (1979) *Cell (Cambridge, Mass.) 17*, 879–887.
Dugaiczyk, A., Woo, S. L. C., Lai, E. C., Mace, M. L., Jr., McReynolds, L., & O'Malley, B. W. (1978) *Nature (London) 274*, 328–333.
Dugaiczyk, A., Woo, S. L. C., Colbert, D. A., Lai, E. C., Mace, M. L., Jr., & O'Malley, B. W. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2253–2257.
Efstratiadis, A., Kafatos, F. C., & Maniatis, T. (1977) *Cell (Cambridge, Mass.) 10*, 571–585.
Gilbert, W., & Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3581–3584.
Ginder, G. D., Wood, W. I., & Felsenfeld, G. (1979) *J. Biol. Chem. 254*, 8099–8102.
Hagenbiichle, O., Santen, M., Steitz, J. A., & Mans, R. J. (1978) *Cell (Cambridge, Mass.) 13*, 551–563.
Hardies, S. C., & Wells, R. D. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 3117–3121.
Hardison, R. C., Butler, E. T., III, Lacy, E., Maniatis, T., Rosenthal, N., & Efstratiadis, A. (1979) *Cell (Cambridge, Mass.) 18*, 1285–1297.
Kafatos, F., Efstratiadis, A., Forget, B., & Weissman, S. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5618–5622.
Konkel, D. A., Maizel, J. V., Jr., & Leder, P. (1979) *Cell (Cambridge, Mass.) 18*, 865–873.
Kozak, M. (1978) *Cell (Cambridge, Mass.) 15*, 1109–1123.
Lacy, E., & Maniatis, T. (1980) *Cell (Cambridge, Mass.) 21*, 545–553.
Lacy, E., Hardison, R. C., Quon, D., & Maniatis, T. (1979) *Cell (Cambridge, Mass.) 18*, 1273–1283.
Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Catterall, J. F., & O'Malley, B. W. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2205–2209.
Lai, E. C., Stein, J. P., Catterall, J. F., Woo, S. L. C., Mace, M. L., Means, A. R., & O'Malley, B. W. (1979) *Cell (Cambridge, Mass.) 18*, 829–842.
Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell (Cambridge, Mass.) 15*, 1157–1174.
Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M., & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 6187–6191.
Mandel, J. L., & Chambon, P. (1980) *Nucleic Acids Res. 7*, 2081–2104.
Mandel, J. L., Breathnach, R., Gerlinger, P., LeMear, M., Gannon, F., & Chambon, P. (1978) *Cell (Cambridge, Mass.) 14*, 641–653.
Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.
Marotta, C. A., Wilson, J. T., Forget, B. G., & Weissman, S. M. (1977) *J. Biol. Chem. 252*, 5040–5053.
Maxam, A., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.
McGhee, J. D., & Ginder, G. D. (1979) *J. Cell Biol. 83*, 31a.
Nishioka, Y., & Leder, P. (1979) *Cell (Cambridge, Mass.) 18*, 875–882.
Proudfoot, N. J. (1977) *Cell (Cambridge, Mass.) 10*, 559–570.
Proudfoot, N. J., & Maniatis, T. (1980) *Cell (Cambridge, Mass.) 21*, 537–544.
Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E., & Goodman, H. M. (1979) *Nucleic Acids Res. 7*, 1137–1146.
Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.
Russell, G. J., Walker, P. M. B., Elton, R. A., & Subak-Sharpe, J. H. (1976) *J. Mol. Biol. 108*, 1–15.
Sanger, F., & Coulson, A. R. (1978) *FEBS Lett. 87*, 107–110.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–518.
Tiemeier, D. C., Tilghman, S. M., Polsky, F. I., Seidman, J. G., Leder, A., Edgell, M. H., & Leder, P. (1978) *Cell (Cambridge, Mass.) 14*, 237–245.
Tilghman, S. M., Tiemeier, D. C., Polsky, F., Edgell, M. H., Seidman, J. G., Leder, A., Enquist, L. W., Norman, B., & Leder, P. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4406–4410.
Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 725–729.
van den Berg, J., van Ooyen, A., Mantei, N., Shambock, A., Grosveld, G., Flavell, R. A., & Weissman, C. (1978) *Nature (London) 276*, 37–44.
Waalmijk, C., & Flavell, R. A. (1978) *Nucleic Acids Res. 5*, 4031–4041.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389–395.
Woo, S. L. C., Dugaiczyk, A., Tsai, M. J., Lai, E. C., Catterall, J. F., & O'Malley, B. W. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3688–3692.