THE STRUCTURE, EXPRESSION AND CHROMOSOMAL ARRANGEMENT OF RABBIT β GLOBIN GENES

Thesis by

Elizabeth Hardy Lacy

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted June 5, 1980)

Acknowledgements

I would first like to thank Richard Axel for encouraging me to join Tom Maniatis' lab as a graduate student. I have learned from many people during the last five years, but I would especially like to acknowledge Tom Maniatis, Brian Seed, Richard Parker and, of course, Ross Hardison. I was supported by a training grant from the National Institutes of Health and received funds for the preparation of this thesis from the Jean Weigle Memorial Fund.

Abstract

Rabbits contain at least three β-like and two α-like globin polypeptides which are differentially synthesized during embryonic development. The structure, expression and chromosomal arrangement of the gene family encoding the β-like polypeptides is described.

A novel procedure was used to isolate the rabbit β globin gene family which does not require the partial purification of single copy genes. Large random rabbit genomic DNA fragments were joined to phage lambda vectors by using synthetic DNA linkers. The resulting recombinant lambda-rabbit DNA molecules were packaged in vitro into viable phage particles and amplified to produce a permanent library of rabbit genomic sequences. The library was screened using cloned globin cDNA probes and an in situ plaque hybridization procedure.

From a screen of the rabbit library, nine clones were isolated which contain four different β-like gene sequences (β1, β2, β3 and β4). Restriction enzyme mapping and blot-hybridization studies indicated that the nine clones contain overlapping restriction fragments, which together encompass 44 kilobase pairs (kb) of contiguous rabbit chromosomal DNA. Therefore, the four rabbit β-like globin genes are physically linked. In addition, all four genes are transcribed off the same strand of DNA in the orientation 5'-β4-β3-β2-β1-3'. A combination of blot-hybridization, R-looping, and DNA sequencing experiments demonstrated the presence of a large intervening sequence in all four genes and a second, smaller intervening sequence in β1, β2 and β4.

Determination of the nucleotide sequence of β1 showed that this gene codes for the second type of two common codominant alleles encoding the rabbit adult β globin chain. A presumptive 1450 nt polyadenylated precursor to β1 mRNA was detected in adult bone marrow. Mature mRNA transcripts from genes β3 and β4 were found in rabbit embryonic erythroid cells, thereby identifying β3 and β4

as embryonic and/or fetal β-like globin genes. RNA blotting experiments confirmed that genes β1, β3 and β4 are differentially expressed during development.

No β2 transcripts were identified in anemic adult bone marrow or reticulocytes or in 12-day rabbit embryos. A DNA sequence analysis demonstrated that β2 cannot code for a functional β globin polypeptide and is therefore a globin pseudogene.

The mRNA expression of β1 was analyzed in a heterologous host cell using DNA-mediated transformations. Thymidine kinase (tk) minus mouse L cells were cotransformed with a lambda clone containing a chromosomal copy of the rabbit adult β globin gene, β1, and the herpes virus tk gene. The tk^+ transformants containing copies of the rabbit gene were analyzed for rabbit β globin transcripts. One transformant was found to contain 5 copies per cell of a cytoplasmic 9S polyadenylated rabbit β globin transcript. Berk-Sharp S1 experiments demonstrated that both intervening sequences are spliced out precisely from the rabbit 9S β globin transcripts in the transformed mouse L cells. This result indicates that RNA splicing mechanisms are not species-specific. However, the 5' termini do not contain 48 ± 5 nucleotides present in the β globin mRNA of rabbit reticulocytes. Therefore, the initiation and/or 5' processing of rabbit β globin nuclear RNA precursors are not identical in transformed mouse fibroblasts and rabbit reticulocytes.

Table of Contents

INTRODUCTION

Hemoglobin is a tetrameric protein composed of two α-like and two β-like polypeptides, each bound with a heme porphyrin ring. In mammals, the α-like and β-like globin chains are encoded by relatively small gene families whose members are differentially expressed during embryonic development and, in some species, during adult erythroid cell maturation as well (Kitchen and Brett, 1974; Bunn, Forget and Ranney, 1977). The subject of the research presented in this thesis is the structure, expression, and chromosomal arrangement of the rabbit β-like globin genes.

Relatively little polypeptide data exist on the number of different β-like and α-like globin chains encoded in the rabbit genome or on the timing of their expression during development. The available experimental results indicate, however, that at least two different embryonic β-like globins, εy and εz, and an embryonic α-like globin, χ, are synthesized in the nucleated erythroid cells derived from the yolk sac blood islands of the embryo (Kitchen and Brett, 1974; Melderis, Steinheider and Ostertag, 1974; Steinheider, Melderis and Ostertag, 1975). As the site of erythropoiesis shifts to the fetal liver, the ε and χ chains are replaced by their adult counterparts, β and α, respectively. A second, additional switch in β globin expression is observed in humans, other primates and hoofed mammals. For example, in humans the embryonic ε globin is replaced in the fetal liver by two nonallelic fetal β-like globins, $^{G}\gamma$ and $^{A}\gamma$, which in turn are replaced shortly before birth by the adult δ and β globins (Bunn et al., 1977). The simpler pattern of β globin gene switching found in rabbits (lagomorphs) is also seen in mice and hamsters (rodents). It should be noted, however, that the absence of fetal-specific β globin polypeptides and thus, the absence of a second switch, has not been definitively demonstrated by experiments in either the rabbit, mouse or hamster β globin gene system.

The globin gene family provides an interesting system for a study of molecular mechanisms involved in differential gene expression during development. Although

the patterns of globin gene switching have been characterized in several mammals, including humans, mice, rabbits, and sheep (Marks and Rifkind, 1972; Benz et al., 1979), rabbits are a preferred organism for the study of differential globin gene expression during early embryogenesis. There are obvious obstacles to obtaining human embryos with a precisely known time of conception. Mice can provide only small amounts of erythroid tissue, especially in the early stages of embryogenesis. In contrast, the time of conception can be controlled relatively easily in rabbits; and each pregnant female carries about ten embryos, which provide sufficient material for biochemical analysis.

The study of globin genes from different species should provide important information regarding the evolution of regulatory mechanisms. An analysis of the phylogeny of globin genes indicates that gene duplication and deletion have occurred independently in the different mammalian α and β globin gene families (Efstratiadis et al., 1980). Furthermore, the observation that distinct embryonic and fetal β-like sequences appeared approximately at the time of mammalian radiation suggests that patterns of globin gene switching during early development evolved independently in the ancestors of the present-day mammalian species. Consequently, it is reasonable to consider the possibility that regulatory mechanisms involved in differential globin gene expression have evolved in a species-specific manner. Therefore, a study of several globin gene families allows for a comparison of different mechanisms used to control the expression of the same functional gene(s).

Chapter 1: Library construction and gene isolation

In order to study globin gene switching at the nucleic acid level in rabbits, we established procedures to isolate the members of the β globin gene family. Briefly, a collection or library of recombinant bacteriophage was constructed by inserting a random population of rabbit DNA fragments 15 to 20 kilobases long into

a phage lambda vector (Blattner et al., 1977). The lambda-rabbit recombinant DNA molecules were packaged into viable phage particles in vitro (Sternberg et al., 1977) and the resulting recombinant phage were grown in host bacteria. The library was screened using cloned globin cDNA probes and an in situ plaque hybridization procedure (Benton and Davis, 1977). The library of rabbit DNA which we constructed contains 7.8×10^5 independent recombinant phage, a number large enough to contain any single copy sequence with a 99% probability (Clarke and Carbon, 1976).

From a screen of 750,000 plaques, nine clones were isolated which contain sequences homologous to rabbit adult β globin mRNA. Restriction mapping and blot-hybridization experiments revealed that four of the clones contain the rabbit adult β globin gene. Furthermore, two of these adult β globin clones contain additional β-like globin gene sequences. This provided the first evidence for gene linkage in the rabbit β globin gene family.

Chapter 2: Linkage arrangement of rabbit β-like globin genes

A total of four different rabbit β-like globin gene sequences, designated β1, β2, β3, and β4, were identified in the set of nine clones isolated from the rabbit DNA library. To establish the linkage arrangement of the four β-like genes, we carried out extensive restriction mapping and hybridization experiments which showed that the set of nine clones contains overlapping restriction fragments encompassing 44 kilobase pairs of contiguous rabbit chromosomal DNA. Therefore, all four rabbit β-like genes are physically linked. In addition, all four genes are transcribed from the same strand of DNA in the orientation 5'-β4-β3-β2-β1-3'. A comparison of the restriction map of β1 in cloned DNA to the map of the adult β globin gene in rabbit chromosomal DNA identified β1 as the adult β globin gene.

Chapter 3: Rabbit β-like globin gene fine structure and expression

In order to study the structure and expression of each of the four rabbit

β-like globin genes, we carried out nucleotide sequencing studies in combination with blot-hybridization and R-looping experiments using RNA isolated from both embryonic and adult erythroid cells. These studies confirmed the identification of β1 as the adult β globin gene and demonstrated the presence of a large intervening sequence in all four genes. A second, smaller intervening sequence was also identified in the β1 and β4 genes. No detectable homology was found among the large intervening sequences in the β1-β4 genes.

RNA blotting experiments detected a polyadenylated transcript of approximately 1450 nucleotides in bone marrow cells of anemic adult rabbits which hybridizes to an intervening sequence probe from β1. Presumably this transcript corresponds to a nuclear RNA precursor containing both the intron and mRNA sequences of the β1 gene. We identified mRNA transcripts from genes β3 and β4 in nucleated reticulocytes from 12-day rabbit embryos. In addition, RNA blotting studies demonstrated that β3 and β4 sequence probes hybridize more efficiently to embryonic erythroid RNA than to adult bone marrow RNA. Therefore, β3 and β4 are embryonic and/or fetal β-like globin genes and β1, β3 and β4 are differentially expressed during development.

Chapter 4: Structure of a β globin pseudogene

The β2 gene was not identified with any known globin polypeptide. In addition, no β2 transcripts were detected in adult bone marrow and reticulocyte RNAs nor in RNA purified from embryonic erythroid cells. To learn whether β2 contains deletions, insertions or base changes that would generate a nonfunctional β globin gene, I determined the nucleotide sequence of this gene and compared the sequence to that of a known functional β globin gene, β1.

α and β globin genes which cannot be identified with known polypeptides have now been observed in several mammalian species, including mouse, human,

and goat. DNA sequence analyses on a mouse α gene (Vanin et al., 1980; Nishioka, Leder and Leder, 1980), a human α gene (Proudfoot and Maniatis, 1980) and a mouse β gene (Jahn et al., 1980) have demonstrated that these sequences are globin pseudogenes, i.e., α- and β-like sequences that do not encode functional α- and β-like globin polypeptides.

A comparison of the protein coding sequences of β1 and β2 revealed that the β2 gene does not encode a functional β globin polypeptide and is therefore a β globin pseudogene. In addition, the alignment of the β1 and β2 nucleotide sequences showed that β2 contains two intervening sequences at the same locations in the globin protein coding sequence as β1 and all other sequenced β globin genes. However, an examination of the DNA sequences at the intron/exon junctions suggested that a β2 precursor mRNA could not be normally spliced. The nucleotide sequence of β2 presented in this paper also made possible a detailed comparison of the flanking and noncoding sequences of β2 and β1 as well as a discussion of the phylogenetic relationship between these two genes.

Chapter 5: The use of DNA-mediated gene transfer to study the in vivo expression of cloned β globin genes

In a collaboration with Richard Axel's lab at Columbia University, we have developed an in vivo system for studying the functional significance of various features of DNA sequence organization. Axel's lab has demonstrated that cellular genes coding for selectable biochemical functions can be stably introduced into cultured mammalian cells by DNA-mediated gene transfer (Wigler et al., 1977, 1978; Pellicer et al., 1978). The biochemical transformants are identified by the stable expression of a gene coding for a selectable marker. Furthermore, these transformants represent a subpopulation of competent cells that integrate other physically unlinked genes for which no selective criteria exist (Wigler et al., 1979). As an initial step

toward studying the transcription and translation of purified eukaryotic genes in a heterologous host, thymidine kinase minus (tk^-) mouse L cells were cotransformed with λRβG1, a lambda clone containing the rabbit adult β globin gene, β1, and the herpes virus tk gene. Six of eight tk^+ transformants were found to contain from 1 to 20 copies of the rabbit β globin gene.

Barbara Wold in Axel's lab studied the expression of the rabbit gene in the six independent transformants. Solution hybridization experiments described in this paper detected transcripts in one of these cell lines at a steady state concentration of 5 copies per cell. An RNA blot analysis showed that these transcripts are polyadenylated and migrate as 9S on methylmercury gels. In order to define the 5' and 3' boundaries of the rabbit globin sequences expressed in the transformed fibroblasts along with the internal processing sites, I used a modification of the Berk-Sharp S1 mapping technique (Berk and Sharp, 1977) which makes it possible to detect less than 5 copies per cell of a mature mRNA transcript. Using this procedure, I determined that the intervening sequences are removed precisely from the rabbit β globin transcripts present in the transformed mouse fibroblasts. This result indicates that RNA splicing mechanisms are not species-specific. However, the initiation or 5' processing of nuclear precursors to rabbit β globin mRNA is not identical in transformed mouse L cells and rabbit reticulocytes since the 5' termini of the 9S rabbit globin transcripts in the mouse cells do not contain 48 $\pm$ 5 nucleotides present in the mature β globin mRNA of rabbit reticulocytes.

REFERENCES

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**, 180-182.

Benz, E. J., Kretschmer, P. J., Geist, C. E., Kantor, J. A., Turner, P. H. and Nienhuis, A. W. (1979). Hemoglobin switching in sheep. J. Biol. Chem. **254**, 6880-6888.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNA by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**, 721-732.

Blattner, F. R., Williams, B. G., Blechl, A. E., Thompson, K. D., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, 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.

Bunn, H. F., Forget, B. G. and Ranney, H. M. (1977). Human Hemoglobins. (Philadelphia: W. B. Saunders Co.)

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. Cell **9**, 91-99.

Efstratiadis, A., et al. (1980). DNA sequence comparison of the human β-like globin genes. Cell, submitted.

Jahn, C. L., Hutchinson, C. A. III, Phillips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F. and Edgell, M. H. (1980). DNA sequence organization of the β-globin complex in the BALB/c mouse. Submitted for publication.

Kitchen, H. and Brett, I. (1974). Embryonic and fetal hemoglobin in animals. Ann. N.Y. Acad. Sci. **241**, 653-670.

Marks, P. A. and Rifkind, R. A. (1972). Protein synthesis: its control in erythropoiesis. Science **175**, 955-961.

Melderis, H., Steinheider, G. and Ostertag, W. (1974). Evidence for a unique kind of α-type globin chain in early mammalian embryos. Nature **250**, 774-776.

Nishioka, Y., Leder, A. and Leder, P. (1980). An unusual alpha globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA, in press.

Proudfoot, N. J. and Maniatis, T. (1980). The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. Cell, submitted.

Pellicer, A., Wigler, M., Axel, R. and Silverstein, S. (1978). The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. Cell **14**, 133-141.

Steinheider, G., Melderis, H. and Ostertag, W. (1975). Embryonic ε chains of mice and rabbits. Nature **257**, 714-716.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). In vitro packaging of a λ Dam vector containing Eco RI DNA fragments of E. coli and phage P1. Gene **1**, 255-280.

Vanin, E. F., Goldberg, G. I., Tucker, P. W. and Smithies, O. (1980). A mouse alpha globin-related pseudogene (ψα30.5) lacking intervening sequences. Nature, in press.

Wigler, M., Pellicer, A., Silverstein, S. and Axel, R. (1978). Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. Cell **14**, 725-731.

Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C. and Axel, R. (1977). Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell **11**, 223-232.

Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell **16**, 777-785.

Chapter 1

Cell, Vol. 15, 687-701, October 1978.

The Isolation of Structural Genes from Libraries of Eucaryotic DNA

Tom Maniatis, Ross C. Hardison, Elizabeth Lacy, Joyce Lauer, Catherine O'Connell and Diana Quon
Division of Biology
California Institute of Technology
Pasadena, California 91125
Gek Kee Sim and Argiris Efstratiadis
Cellular and Developmental Biology
Harvard University Biological Laboratories
Cambridge, Massachusetts 02138

Summary

We present a procedure for eucaryotic structural gene isolation which involves the construction and screening of cloned libraries of genomic DNA. Large random DNA fragments are joined to phage lambda vectors by using synthetic DNA linkers. The recombinant molecules are packaged into viable phage particles in vitro and amplified to establish a permanent library. We isolated structural genes together with their associated sequences from three libraries constructed from Drosophila, silkmoth and rabbit genomic DNA. In particular, we obtained a large number of phage recombinants bearing the chorion gene sequence from the silkmoth library and several independent clones of β-globin genes from the rabbit library. Restriction mapping and hybridization studies reveal the presence of closely linked β-globin genes.

Introduction

With few exceptions, it has not been possible to purify single-copy structural genes from complex eucaryotic genomes using conventional biochemical methods. The chromosomal arrangement of such genes, however, and the nature of their associated sequences, can be studied using recombinant DNA techniques. If random fragments of genomic DNA are cloned ("cloned library") and the number of clones is large enough for complete sequence representation, in principle any gene can be isolated by screening the library with a specific hybridization probe (for example, see Young and Hogness, 1977). The feasibility of constructing libraries from the DNA of organisms with small genomes such as Drosophila or yeast has been demonstrated (Wensink et al., 1974; Carbon et al., 1977). Construction of libraries from larger genomes, however, has not been attempted because of technical limitations, especially those related to screening large numbers of clones. As an alternative, methods have been developed to substantially enrich for structural gene sequences from mammalian DNA prior to cloning (Tilghman et al., 1977, 1978; Tonegawa et al., 1977).

Three recent technical advances, all concerning phage lambda, have made possible the construction and screening of DNA libraries from complex genomes. First, a rapid in situ plaque hybridization procedure has been developed (Benton and Davis, 1977). Second, EK2-certified lambda cloning vectors have been constructed (Leder, Tiemeier and Enquist, 1977; Blattner et al., 1977). Third, in vitro packaging systems have been developed for use in recombinant DNA experiments (Hohn and Murray, 1977; Sternberg, Tiemeier and Enquist, 1977). In vitro packaging substantially increases the efficiency of introducing lambda DNA into bacterial cells compared to transfection methods.

The strategy which we used to construct libraries of eucaryotic DNA is outlined in Figure 1. In brief, high molecular weight DNA is fragmented either by shearing followed by S1 nuclease treatment, or by a nonlimit endonuclease digestion with restriction enzymes generating molecules with blunt ends. Molecules of approximately 20 kb are selected by preparative sucrose gradient centrifugation and rendered Eco RI-resistant by treatment with Eco RI methylase. Synthetic DNA linkers bearing Eco RI recognition sites are covalently attached by blunt-end ligation using T4 ligase. Cohesive ends are then generated by digestion with Eco RI. These molecules are covalently joined to phage λ DNA, and the hybrid DNA molecules are then packaged into viable phage particles in vitro. Amplification of these phage produces a permanent library of eucaryotic DNA sequences. To demonstrate the generality of the approach, we present the characterization of libraries constructed from the DNA of three organisms: Drosophila, silkmoth (Antheraea polyphemus) and rabbit. The phage λ vectors used for the construction of these libraries (referred to as "Charon 4" in the text) were the EK1 vector Charon 4 (Drosophila and silkmoth libraries) and its derivative EK2 vector Charon 4A (rabbit) (for a description of these vectors, see Blattner et al., 1977).

This paper demonstrates that these libraries can be rapidly screened for individual structural gene sequences. In particular, we describe the isolation and preliminary characterization of cloned rabbit β-globin genes. Each step of the approach outlined in Figure 1 is described in detail below.

Results

Preparation of 20 kb Eucaryotic DNA Fragments

The internal Eco RI fragments of Charon 4 can be replaced by DNA inserts (in the size range of 8.2-22.2 kb) bearing Eco RI cohesive ends (Blattner et

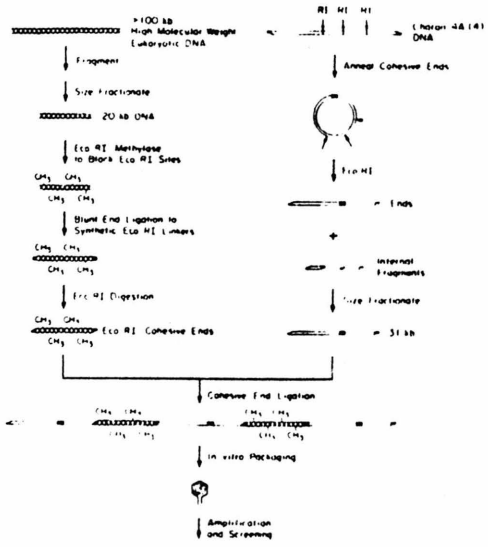

Figure 1. Schematic Diagram Illustrating the Strategy Used to Construct Libraries of Random Eucaryotic DNA Fragments

al., 1977). For this purpose, high molecular weight eucaryotic DNA was fragmented and fractionated by size to obtain molecules of approximately 20 kb. This reduces the number of plaques required for screening an entire genome and minimizes the possibility of small eucaryotic DNA fragments joining to each other and subsequently to the vector. Although a nonlimit Eco RI digestion is the most direct method of obtaining large DNA fragments with Eco RI cohesive ends (Glover et al., 1975), we were concerned that a nonrandom distribution of Eco RI sites within and adjacent to particular structural genes might result in their selective loss from the population of size-fractionated DNA. We therefore used two methods to obtain random eucaryotic DNA fragments, both of which generate molecules with blunt ends that can be joined to Eco RI linkers and subsequently inserted into the cloning vector.

In one method, DNA was fragmented by shearing and the ends were trimmed using S1 nuclease. Unfortunately, this method of preparing blunt-ended DNA fragments is rather inefficient (see also Scheller et al., 1977b; Seeburg et al., 1977). An alternative, and in fact more efficient, approach is to perform a nonlimit restriction endonuclease digestion with two enzymes that cleave frequently and generate blunt-ended molecules – that is, Hae III (GGCC) and Alu I (AGCT) (Roberts, 1976). The combined activities of these two enzymes under conditions of nonlimit digestion should generate a collection of large fragments approaching the random sequence representation of sheared DNA more closely than the products of a nonlimit Eco RI digestion; the reason for this is that a recognition site for each of the two enzymes on DNA should occur once every 256 (4^4) nucleotides, whereas Eco RI recognition sites should be present an average of once every 4096 (4^6) nucleotides (uncorrected for base composition). The greater the number of possible cleavage sites, the larger the number of possible ways of generating a 20 kb fragment by nonlimit digestion of a higher molecular weight molecule, and thus the more random the collection of resulting fragments.

The results of an electrophoretic analysis of nonlimit Hae III and Alu I digests of rabbit DNA are shown in Figure 2. Three reactions with different enzyme to DNA ratios were performed separately for each enzyme, and the digestion products containing a substantial fraction of 20 kb fragments were pooled and fractionated on a 10–40% sucrose gradient.

Modification of Eco RI Sites in Eucaryotic DNA

Since the synthetic Eco RI linkers attached to eucaryotic DNA must be cleaved by Eco RI to generate cohesive ends, the Eco RI sites within the DNA fragments must first be rendered resistant to cleavage. This was accomplished by reacting the eucaryotic DNA with the Eco RI modification-methylase in the presence of S-adenosyl-L-methionine (Greene et al., 1975). Small aliquots of the methylation reaction mixture were taken before and after the addition of methylase and mixed with λ DNA to monitor the extent of methylation; following electrophoresis, the appearance of discrete λ DNA bands reveals incomplete methylation. The results of a typical assay are shown in Figure 3. The mixture of λ and eucaryotic DNAs taken before the addition of methylase is digested to completion, while the methylated DNA is totally resistant to Eco RI cleavage.

Joining of Synthetic DNA Linkers to Eucaryotic DNA

Duplex DNA linker molecules bearing restriction endonuclease recognition sites have been chemically synthesized (Bahl et al., 1977; Scheller et al., 1977a) and used to insert a number of different DNA molecules into plasmids (Heyneker et al., 1976; Scheller et al., 1977b; Shine et al., 1977; Ullrich et al., 1977). In each case, the linkers were first blunt-end ligated to the DNA of interest and then digested with the appropriate restriction enzyme to generate molecules with cohesive termini that could be joined to a vector with complementary ends.

We attached Eco RI dodecameric linkers to in vitro methylated eucaryotic DNA using T4 ligase and assayed for blunt-end ligation by the formation

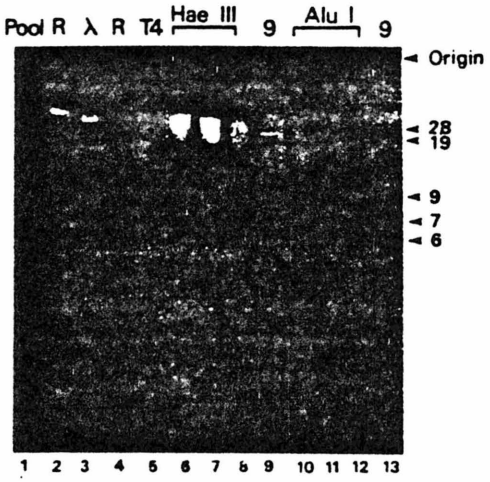

Figure 2. Agarose Gel Electrophoresis of Nonlimit Hae III and Alu I Digests of Rabbit Liver DNA

High molecular weight rabbit DNA was digested with different amounts of Hae III or Alu I and electrophoresed on a horizontal 0.5% agarose gel containing 0.5 μg/ml ethidium bromide, and the DNA was photographed on a short wave ultraviolet transilluminator (Sharp et al., 1973). (1, Pool) the pooled DNA from the Hae III and Alu I digests of slots 6–8 and 10–12; (2 and 4, R) undigested rabbit liver DNA; (3, λ) undigested wild-type λ DNA (49.4 kb) (Blattner et al., 1977); (5, T4) undigested T4 DNA (171 kb) (Kim and Davidson, 1974); (6–8, Hae III) Hae III-digested rabbit liver DNA; (9 and 13, 9) size markers, Eco RI-digested Charon 9 DNA; (10–12, Alu I) Alu I-digested rabbit DNA. The approximate sizes of the markers in kb are indicated on the side of the figure.

of linker oligomers. After ligation, the DNA was sedimented through a 10–40% sucrose gradient (or passed over a Sepharose 2B column) to remove unincorporated linker oligomers. Without this step, the digestion of linkers attached to eucaryotic DNA is difficult, since linker oligomers not incorporated into eucaryotic DNA compete for Eco RI. Eco RI digestion of the DNA thus purified provides large fragments with Eco RI cohesive ends that can be joined to the vector DNA.

Ligation of Eucaryotic DNA to Charon 4 DNA

Charon 4 contains three Eco RI cleavage sites (Figure 1). Digestion with Eco RI produces two internal fragments with genes nonessential for phage viability and two end fragments. To maximize the efficiency of in vitro recombination and to minimize the number of nonrecombinant phage in the library, we removed the internal fragments by sucrose gradient centrifugation. After annealing the cohesive ends of Charon 4 DNA and Eco RI digestion, excellent separation of the cohered λ DNA arms (31 kb) and the internal fragments (7 and 8 kb) can be achieved. To determine the ratio of vector DNA to eucaryotic DNA that produces the

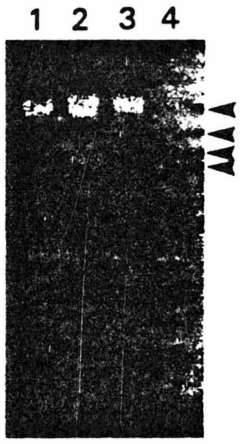

Figure 3. Assay for Methylation of Eco RI Sites in Eucaryotic DNA

A mixture of eucaryotic DNA and Charon 4A DNA was reacted with Eco RI methylase as described in Experimental Procedures and analyzed by agarose gel electrophoresis. (1) DNA mixture plus Eco RI methylase, minus Eco RI nuclease. (2) DNA mixture plus Eco RI methylase, plus Eco RI nuclease. (3) DNA mixture minus Eco RI methylase, minus Eco RI nuclease. (4) DNA mixture minus Eco RI methylase, plus Eco RI nuclease. The arrows indicate the positions of phage λ Eco RI fragments.

smallest number of background plaques without reducing the absolute yield of recombinants, we ligated varying amounts of eucaryotic DNA with a constant amount of purified λ DNA arms, packaged the DNA into phage and determined the number of plaque-forming units (pfu). Using the established optimal ratio, ligation reactions were performed at high DNA concentrations to minimize intramolecular joining and to maximize the formation of concatemeric DNA recombinants, the substrate for in vitro DNA packaging.

In Vitro Packaging of DNA into Phage Particles

The number of independently derived phage recombinants (library size) required for a 99% probability of finding any given single-copy sequence in the library can be calculated if the average size of the eucaryotic DNA inserts is known (Clarke and Carbon, 1976). Thus 7×10^8 recombinants are required for a mammalian DNA library of 20 kb DNA inserts. Using the $CaCl_2$ transfection procedure of Mandel and Higa (1970), approximately $2\text{–}10 \times 10^3$ pfu/μg of cleaved and religated λ vector DNA are obtained (in contrast to 10^6 pfu/μg of intact λ DNA). Thus approximately 100–400 μg of DNA fragments attached to synthetic linkers are needed for the construction of a complete library, an amount difficult to obtain. Fortunately, efficiencies of 2 and 0.15×10^7 pfu/μg have been achieved for intact and religated λ DNAs, respectively, using in

vitro packaging procedures (Hohn and Murray, 1977; Sternberg et al., 1977).

To use this technique for mammalian DNA cloning, it was necessary to demonstrate that the procedure does not alter the biological containment features of Charon 4A. Most in vitro packaging procedures involve the temperature induction of λ lysogens which carry amber mutations in different genes required for packaging. Each lysogen alone is incapable of producing viable phage particles, but mixed lysates of the two strains complement in vitro to convert λ DNA into a plaque-forming particle. If hybrid phage DNA carrying eucaryotic sequences is added to a mixed extract, there is a possibility that endogenous prophage DNA will recombine with the hybrid DNA in vitro or during subsequent in vivo amplification to produce DNA carrying the wild-type markers of the prophage.

Sternberg et al. (1977) have developed a packaging system that minimizes these problems. First, the prophage of their strains carry the λ*b2* mutation, which removes part of the attachment site and therefore prevents prophage excision after induction. Second, the lysogens are recombination-deficient (the prophage is red and the host is recA). To reduce further the chance of prophage DNA packaging and recombination we ultraviolet-irradiated the cells prior to their use in in vitro packaging reactions. Hohn and Murray (1977) found that both recombination and prophage packaging in their extracts could be suppressed by irradiation with ultraviolet light. To obtain EK2 certification for the system of Sternberg et al. (1977), we examined extracts derived from ultraviolet-irradiated cells for recombination and for the presence of in vitro packaged prophage. A procedure for the preparation and testing of in vitro packaging extracts for EK2 experiments is presented in Experimental Procedures.

Following in vitro packaging of recombinant DNA, the resulting phage particles were separated from cellular debris by sedimentation on a CsCl step gradient. This procedure concentrates the phage and removes material present in the extracts which inhibits the growth of bacterial cells.

Amplification of Libraries

An essential feature of our strategy for gene isolation is to establish a permanent library that can be repeatedly screened. To achieve this, it is necessary to amplify the in vitro packaged recombinant phage and to store the library in the form of a plate lysate. There is, of course, a risk that a particular recombinant phage will exhibit a growth disadvantage and will be eliminated from the library during amplification. It is therefore important to minimize competitive growth. We accomplished this by plating the in vitro packaged phage on agar plates at low density (10,000 pfu per 15 cm diameter plate). In this manner, the phage are amplified approximately 0.1 to 1 × 10^9 fold and recovered as a plate lysate.

Efficiency of Cloning Eucaryotic DNA

The efficiency of eucaryotic DNA cloning under our conditions depends primarily upon the quality of the in vitro packaging extracts (which varies between preparations) and the fraction of DNA fragments bearing Eco RI linkers. The latter is determined by the fraction of molecules with two blunt ends which, in turn, depends upon the method of preparation. The efficiency of our extracts prepared for EK1 experiments varies from 2–20 × 10^7 pfu/μg of intact λ DNA. Extracts prepared for EK2 experiments are consistently less efficient, varying from 0.4–5 × 10^7 pfu/μg of intact λ DNA. When using cleaved and religated DNA, this efficiency drops. Even the lowest efficiency observed (3.8 × 10^4 pfu/μg of DNA in the case of the rabbit library), however, is higher than that reported for cloning religated λ DNA by transfection (2–10 × 10^3 pfu/μg; Thomas, Cameron and Davis, 1974; Hohn and Murray, 1977).

Analysis of the data of Table 1, which describes the characterization of the libraries, reveals that DNA fragments produced by nonlimit endonuclease digestion (rabbit library) are cloned more efficiently than those produced by shearing followed by S1 nuclease treatment (Drosophila and silkmoth libraries). When the number of plaques formed per μg of eucaryotic DNA is normalized to the particular efficiency of the in vitro packaging extract used to construct each library, it becomes evident that the rabbit DNA was cloned 15.8 and 3.4 times more efficiently than the Drosophila and silkmoth DNAs, respectively. (We cannot, of course, rule out the unlikely possibility that the cloning efficiency is genome-specific).

Fraction of Clones Containing Eucaryotic DNA

To test whether a library containing the entire complement of genomic DNA can be constructed, we have measured the number of recombinants carrying eucaryotic sequences, the average size of eucaryotic DNA inserts and the single-copy DNA sequence representation in one of the libraries (Drosophila).

To estimate the number of recombinant phages in each library, it was necessary to determine the number of background (nonrecombinant) phage present. The number of background plaques was minimized by separating the annealed cohesive ends from the internal λ DNA fragments. When the purified cohesive ends of Charon 4 were ligated without the addition of eucaryotic DNA, however, a small number of pfu (on the order of a few percent

Table 1. Characterization of Libraries

Library	1 Efficiency of Extract	2 Plaques per μg of Eucaryotic DNA	3 Relative Efficiency of Cloning[a]	4 % Blue Plaques after Amplification	5 Total Number of Independent Recombinant Phage Recovered	6 Mean Length of Eucaryotic DNA	7 Number of Recombinant Phage Required for a "Complete" Library[b]
Drosophila	1×10^8	6.0×10^4	1	3.0	6.0×10^6	16 kb	4.8×10^4
Silkmoth	2×10^8	5.6×10^5	4.6	7.0	2.8×10^6	19 kb	2.4×10^5
Rabbit	4×10^8	3.8×10^4	15.8	2.6	7.8×10^6	17 kb	8.1×10^5

[a] This number was determined by dividing the number of column 2 by that in column 1 and normalizing to the value calculated for the Drosophila library.

[b] Calculated as described by Clarke and Carbon (1976) using the values in column 6 and assuming genome sizes of 1.65×10^9 bp for Drosophila (Rudkin, 1972), and 1×10^9 and 3×10^9 bp for silkmoth and rabbit, respectively (J. Yeh, L. Villa-Komaroff and A. Efstratiadis, unpublished results).

of the final number of plaques in the libraries) was recovered from the in vitro packaging reaction. The number of nonrecombinant phage in the libraries can also be estimated by an indicator plate assay. One of the internal Charon 4 fragments carries the E. coli lactose operator-promoter region and the gene for β-galactosidase. The presence of this fragment in library phage DNA can be detected by plating the phage on a lawn of lac⁻ E. coli grown on an Xgal indicator plate. Phage carrying an intact β-galactosidase gene produce blue plaques under these conditions (for a discussion of this assay, see Blattner et al., 1977). The number of blue (nonrecombinant) plaques observed after amplification is small (Table 1), indicating that most of the library phage carry eucaryotic DNA.

An independent, nonquantitative estimate of the degree of nonrecombinant phage contamination of the libraries can be obtained by determining the amount of internal Charon 4 DNA fragments in library DNA. This was accomplished by growing an aliquot of each library in liquid culture, purifying recombinant phage DNA and digesting the DNA with Eco RI. Figure 4 (lanes 2 and 4) shows the results of such an analysis of Drosophila and rabbit library DNAs. In addition to the left and right arms of the Charon 4 DNA, a characteristic smear of restriction endonuclease-digested eucaryotic DNA can be observed in the two library DNAs (compare to lanes 1 and 5 of Figure 4). The gel was intentionally overloaded to show that a small number of contaminating internal phage DNA fragments are present in both libraries. A similar result was obtained with the silkmoth library (not shown). All our assays together clearly demonstrate that most of the phage in each library contain eucaryotic DNA.

Mean Length of Eucaryotic DNA in the Libraries

The number of independent phage recombinants required for a complete library depends upon the average size of the cloned eucaryotic DNA inserts, which we estimated for each of the three libraries by CsCl sedimentation equilibrium analysis. Since the amount of protein in different λ phage is constant and the buoyant density depends upon the DNA/protein ratio, the distribution of DNA sizes in a λ phage population can be determined by measuring the distribution of phage in a CsCl density gradient (Weigle, Meselson and Paigen, 1959; Davidson and Szybalsky, 1971; Bellet, Busse and Baldwin, 1971). Figure 5 shows the results of a CsCl density gradient analysis of rabbit library phage. The density of each fraction was determined by its position in the gradient relative to two λ phage density markers. The average size of the rabbit DNA inserts calculated from the midpoint of the curve of Figure 5 is 17 kb, resulting in recombinant Charon DNA molecules whose average size is 97% of that of wild-type λ DNA. Similar analyses of the Drosophila (W. Bender and D. S. Hogness, personal communication) and silkmoth libraries yielded insert sizes of 16 and 19 kb, respectively.

Knowing the approximate size of the eucaryotic DNA fragments carried in each library and the complexity of each haploid genome (Table 1), we can calculate the number of independent recombinant phage needed to find any given single-copy sequence in the library with a probability of 0.99 (Clarke and Carbon, 1976), assuming that the entire genome consists of single-copy DNA sequences. Thus a 99% complete library of Drosophila, silkmoth or rabbit DNA would consist of 4.8×10^4, 2.4×10^5 or 8.1×10^5 recombinant phage, respectively. By comparing the theoretical library sizes to the actual ones (Table 1), we conclude that our libraries are "complete."

Sequence Representation in Library DNA

To determine whether a significant fraction of single-copy sequences is lost during amplification of

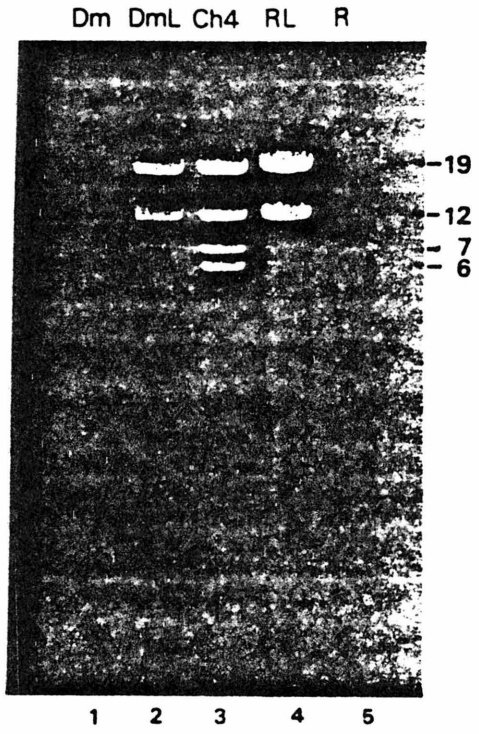

Figure 4. Presence of Internal Charon 4 DNA Fragments in the Drosophila and Rabbit Libraries

Aliquots of the Drosophila and rabbit libraries were each grown in liquid culture, and DNA was isolated as described in Experimental Procedures. The DNA samples were digested with Eco RI and analyzed on a 0.5% agarose gel. (1, Dm) Drosophila DNA, (2, DmL) Drosophila library DNA, (3, Ch4) Charon 4 DNA, (4, RL) rabbit library DNA, (5, R) rabbit liver DNA.

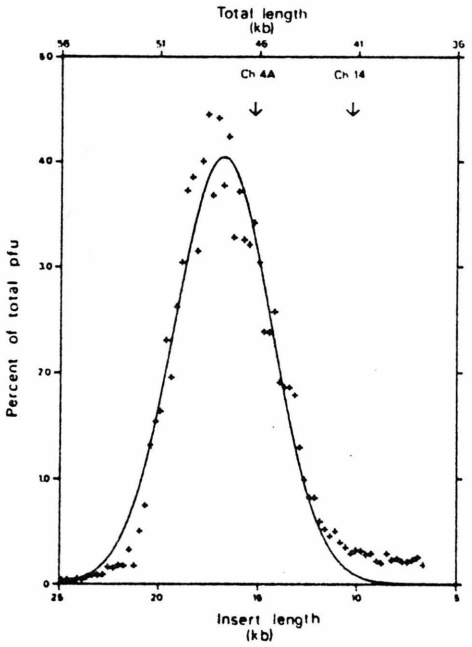

Figure 5. Average Size of Eucaryotic DNA Fragments in the Rabbit Library

The density distribution of phage from the rabbit library was determined by CsCl equilibrium centrifugation. Gradient fractions were titered, and the distribution of library phage (+) was computer fitted to a Gaussian curve (solid line). The arrows indicate the positions of the Charon 4A (93.6% λ) and Charon 14 (83.7% λ) phage density markers. Phage DNA length (upper abscissa) was calculated from the density relative to marker phage (Davidson and Szybalski, 1971). The insert length (lower abscissa) was determined by subtracting the length of the Charon 4A DNA arms (31 kb) from the total length of the phage DNA. The half-maximal band width of the curve is significantly greater than that observed for the marker phage (data not shown), indicating that a heterogeneous size population of rabbit DNA inserts is present in the library. This heterogeneity is not predicted by the relationship between insert size and packaging efficiency reported by Sternberg et al. (1977).

the libraries, we measured the single-copy complexity of library DNA using the procedure of Galau et al. (1976). Tritium-labeled single-copy tracer DNA was prepared from Drosophila DNA and driven with both sheared library and embryo DNAs. As shown in Figure 6, the reassociation rate and extent of reaction of the tracer are identical in both cases. We conclude that within the sensitivity of our measurements, the Drosophila library contains the entire complement of single-copy sequences present in genomic DNA. Phage libraries produced by nonlimit Eco RI digestion of sea urchin DNA (D. Anderson and E. Davidson, personal communication) and Drosophila DNA (R. Robinson and N. Davidson, personal communication) have been shown to be nearly complete by this criterion.

Screening Libraries for Structural Gene Sequences

Assuming an average size of 17 kb for the inserts in the rabbit library and a genome size of 3×10^9 bp, only 1 in 180,000 plaques will carry a particular single-copy sequence. Optimal screening conditions are therefore necessary to identify clones carrying such a sequence. We examined a number of variables in the plaque hybridization procedure of Benton and Davis (1977), including the type of medium used in the agar plates (L broth or NZCY), the strain of host bacteria (KH802 or DP50SupF), the concentration of plating bacteria and the method of preparing filters. Most of the variables had little effect on the intensity or the number of positive signals observed. The most significant differences resulted from varying the concentration of plating bacteria; the best signals were obtained

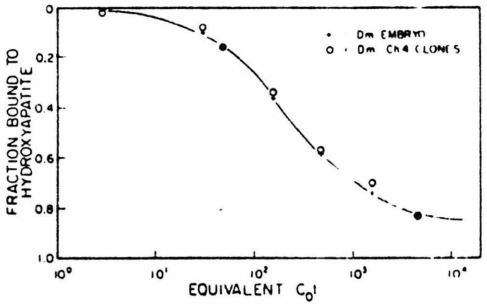

Figure 6. Single-Copy Sequence Representation in the Drosophila Library

A comparison of the reassociation kinetics of Drosophila embryo and Drosophila library DNAs with ^{3}H-labeled single-copy tracer (1 × 10^6 cpm/μg). Drosophila library DNA was prepared from phage grown in liquid culture (5 × 10^8 fold amplification). Single-copy tracer DNA was prepared from Drosophila pupae DNA, and reacted with sheared Drosophila embryo DNA and Drosophila library DNA according to the procedures of Galau et al. (1976). For each point, 3 μg of embryo DNA or 7.5 μg of library DNA were reacted with approximately 800 cpm of single-copy tracer. The library DNA was at 2.5 times the concentration of the embryo DNA on the assumption that 40% of each hybrid phage molecule consists of eucaryotic DNA. The solid line is a computer fit of the embryo data describing a single second-order component.

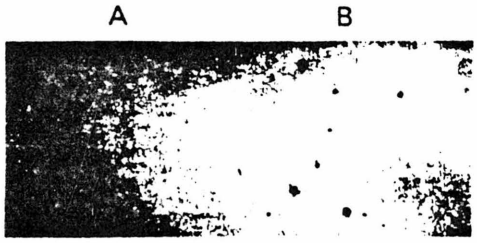

Figure 7. Screening of Silkmoth Library for Chorion Gene Sequences

(A and B) Autoradiograms of duplicate nitrocellulose filters showing specific hybridization of chorion cDNA to silkmoth library phage clones. 5000 phage were plated onto a 10 cm petri dish and incubated for 16 hr at 37°C, and the phage DNA was transferred to two nitrocellulose filters applied in succession. Agarose rather than agar was used in the 0.7% top agar layer. The filters were prepared for hybridization as described in Experimental Procedures and hybridized for 48 hr to 50 ng/ml ^{32}P-cDNA (spec. act. 2.5 × 10^7 cpm/μg) prepared from total chorion mRNA. Filters were washed, dried and exposed to X-ray film for 48 hr using a single intensifier screen. Five strong and two weak positive signals appear on both filters identifying phage clones bearing chorion gene sequences.

using 3.1 × 10^8 exponentially growing bacteria per 15 cm plate. The best correspondence between positive signals on duplicate filters occurred when the filters were sequentially applied to the plates; stacking filters often resulted in failure to observe duplicate signals.

To determine the optimal plaque density for screening, we plated various amounts of a mixture of Charon 4A and a Charon 16-pβG1 hybrid (pβG1 is a β-globin cDNA plasmid; Maniatis et al., 1976) at a constant ratio of 250:1. As many as 20,000 pfu per 15 cm plate could be screened for globin with no apparent loss of signal on autoradiograms. The most serious problem encountered in screening libraries is nonspecific background hybridization to nitrocellulose filters. Using conditions adapted from those developed for Southern transfer experiments (Jeffreys and Flavell, 1977a), however, we reproducibly observed low background.

All three libraries have been successfully screened using gene-specific hybridization probes. Several different phage recombinants that hybridize to different Drosophila cDNA plasmid clones have been selected from the Drosophila library (W. Bender and D. S. Hogness, personal communication). The frequencies at which these clones were detected and the single-copy complexity measurement of Figure 6 indicate that most if not all Drosophila structural gene sequences are present in the library.

The silkmoth library was screened for genes, sequentially expressed during oogenesis, which encode the approximately 100 eggshell (chorion) proteins of the developing oocyte (for a review of this system, see Kafatos et al., 1978). Although the exact number of chorion genes is not known, preliminary evidence suggests that each gene cannot be present in more than a few copies (J. Yeh, W. C. Jones and A. Efstratiadis, unpublished results). With an average insert size of 19 kb in the silkmoth library and a genome size of 10^9 bp, we expected to observe one positive signal per 530 plaques using cDNA transcribed from total chorion mRNA as probe, assuming that the chorion genes are unique. If some of the genes are closely linked, positive signals will be less frequent. Figure 7 (A and B) shows an autoradiogram of duplicate filters prepared from a plate containing 5000 plaques. 100% of the duplicate positive signals proved real when individual plaques were picked onto a bacterial lawn and rescreened. When only single (instead of duplicate) filters were prepared, 88% of the initial positives proved real upon rescreening. To date, screening of 350,000 plaques has yielded 350 independent isolates, about 53% of the number expected if the genes are unlinked.

The rabbit library was screened for globin sequences using cDNA prepared from total globin mRNA. Figure 8A shows an autoradiogram of a filter prepared from a 15 cm agar plate carrying 10,000 plaques. In the example shown, two positives were observed on one filter (a rare event), demonstrating two types of signals. One signal reflects the plaque morphology, while the other contains a head and a comet-like tail. The latter frequently observed morphology could result from

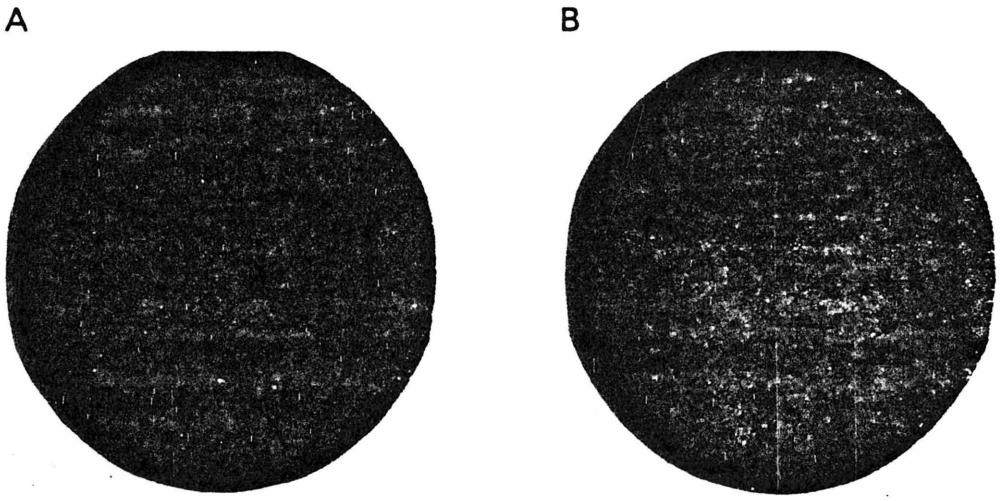

Figure 8. Screening the Rabbit Library for Globin Sequences

(A) Autoradiogram of a nitrocellulose filter prepared from a 15 cm plate containing 10,000 recombinant plaques. The filter was hybridized to 2.4 ng/ml ^{32}P-globin cDNA (5 × 10^8 cpm/μg) for 36 hr, washed as described in Experimental Procedures, dried and exposed to preflashed X-ray film for 48 hr at −70°C using a single intensifier screen. Arrows indicate the locations of two positive signals.

(B) Autoradiogram of a filter prepared from a plate containing 500 plaques obtained by plating a number of plaques from the area on the plate (A) corresponding to the location of one of the two positive signals shown in (A). The filter was hybridized with 25 ng/ml of nick-translated pβG1 DNA (5 × 10^7 cpm/μg) for 12 hr, and exposed to preflashed X-ray film for 24 hr at −70°C using two intensifier screens.

the spreading of phage DNA from the plaque during filter application or subsequent handling. Both spots were shown to be true positives by their appearance on duplicate filters and by rescreening (Figure 8B).

A total of four independent β-globin clones were recovered from 750, 000 plaques screened with globin cDNA. To identify clones carrying β-globin sequences, each clone was hybridized to in vitro labeled pβG1 DNA, a rabbit β-globin cDNA plasmid (Maniatis et al., 1976). With a genome size of 3 × 10^9 bp and cloned inserts of approximately 17 kb, we expected to recover four to five clones of the adult β-globin sequence from 750,000 plaques, a number close to that actually recovered. In this calculation, we assumed that cross-hybridization of adult β-globin probe to embryonic β-like rabbit globin genes is inadequate to allow detection of these genes in a total genome screen.

Characterization of Clones That Hybridize to Rabbit Globin Probes

To demonstrate that the four clones hybridizing to the β-globin plasmid actually carry the β-globin gene sequence, we digested DNA from each clone with Eco RI, fractionated the products on a 1.4% agarose gel, transferred the DNA to a nitrocellulose filter and hybridized them to in vitro labeled globin cDNA. Figure 9A shows the Eco RI cleavage pattern of DNAs from the four β-globin clones. As expected for DNA fragments generated by random cleavage, each clone contains common and unique Eco RI fragments. Presumably the common set contains the β-globin gene and its adjacent sequences, while the unique fragments lie further from the gene in the 5′ or 3′ direction. Figure 9B shows the resulting autoradiogram of the hybridization experiment. In all four clones, two fragments of approximately 2600 and 800 bp hybridize to the probe. These fragments are, respectively, the sizes of the 5′ and 3′ β-globin Eco RI fragments found in genomic DNA (Jeffreys and Flavell, 1977a, 1977b). The identification of these bands is confirmed by detailed restriction mapping and DNA sequence analysis of the cloned DNA (our unpublished results). In addition to these two fragments, RβG2 (lane 2) contains one Eco RI fragment and RβG5 (lane 4) contains three fragments that hybridize weakly to globin cDNA. These data and the failure of these two clones to hybridize α-globin probe indicate that the additional Eco RI fragments correspond to β-like sequences closely linked to the adult β-globin gene.

The 6.3 kb Eco RI fragment of RβG2 and RβG5, which hybridizes weakly compared with the 2.6 kb fragment, may correspond to the faint 6.9 kb Eco RI fragment detected in genomic DNA (Jeffreys and Flavell, 1977a). Preliminary restriction mapping data from RβG2 indicate that the β-like sequence

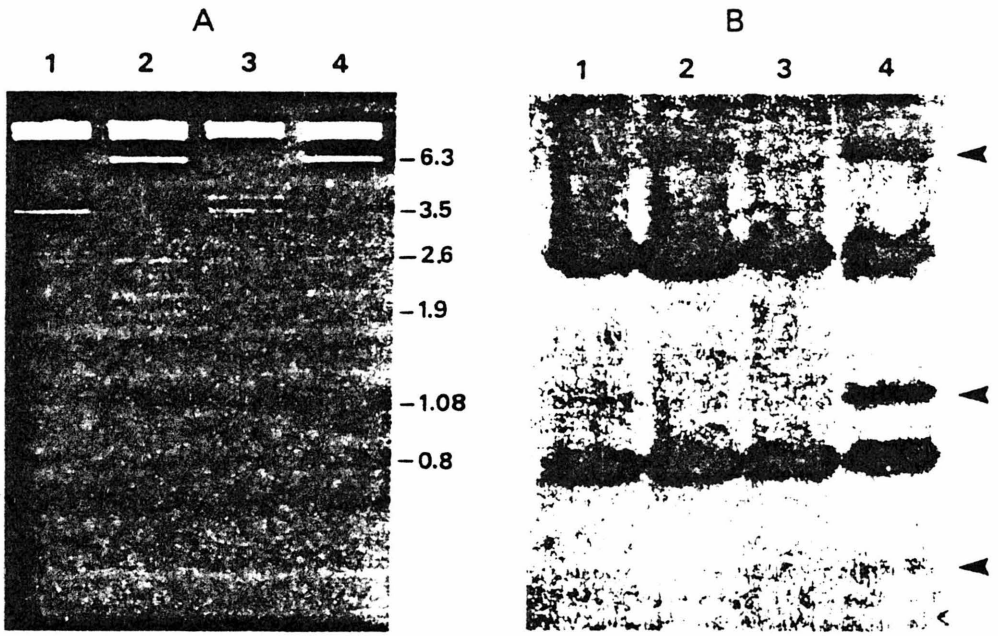

Figure 9. Eco RI Cleavage Patterns of DNAs from Rabbit β-Globin Phage Clones

DNAs from rabbit β-globin phage clones were digested with Eco RI, fractionated on a 1.4% agarose gel, transferred to a nitrocellulose filter (Southern, 1975) and hybridized to ^{32}P-globin cDNA (1 × 10^6 cpm/μg)

(A) Ethidium bromide-stained gel. (A) λCH4A-RβG1; (2) λCh4A-RβG2, (3) λCh4A-RβG3; (4) λCh4A-RβG5

(B) Autoradiogram of the nitrocellulose filter prepared from the gel shown in (A). The sizes of various Eco RI fragments are indicated in kb. The arrows indicate DNA fragments bearing β-like globin sequences that are not a part of the adult β-globin gene.

carried on the 6.3 kb fragment is approximately 9.1 kb away from the adult β-globin gene in the 5′ direction (data not shown). The β-like sequences in RβG5 have not yet been mapped.

Discussion

This paper shows that it is possible to isolate structural genes directly from large eucaryotic genomes by screening libraries of DNA fragments cloned in phage λ. The overall efficiency of the procedure described yields a collection of recombinants large enough to represent the entire genome of a mammalian cell. Such a collection can be amplified by a factor of 10^6, with no apparent loss of sequence complexity, to produce a library of eucaryotic DNA that can be screened repeatedly using different probes. This rapid method of gene isolation provides many advantages over existing techniques. For example, all the members of a family of evolutionarily or developmentally related genes can be isolated in a single step by screening a library with a mixed probe. Furthermore, isolation of a set of overlapping clones, all of which contain a given gene, permits the study of sequences extending many kilobases from the gene in the 5′ and 3′ directions. Moreover, even more distant regions along the chromosome can be obtained by rescreening the library using terminal fragments of the initially selected clones, allowing the isolation of linked genes.

The power of this approach is clearly illustrated by the isolation of globin and chorion genes. The mammalian globin genes constitute a relatively simple family comprised of at least two α-like and at least four β-like embryonic and adult genes, which are expressed during erythropoiesis in different cell populations at different developmental times (Clissold, Arnstein and Chesterton, 1974; Melderis, Steinheider and Ostertag, 1974; Steinheider, Melderis and Ostertag, 1975, 1977). The gene family that codes for the chorion proteins of the silkmoth Antheraea polyphemus is considerably more complex (Kafatos et al., 1978). Approximately 100 different chorion genes are sequentially expressed during oogenesis. We have used the procedure of cDNA cloning to purify to homogeneity sequences that correspond to individual members of the chorion gene family (Maniatis et al., 1977; Sim et al., 1978; G. K. Sim, unpublished results).

Some of these cDNA plasmids have been classified according to the time of their expression during development (Kafatos et al., 1978). Using these cDNA plasmids to rescreen the phage recombinants that hybridize to total chorion cDNA, it will be possible to isolate and study genes that are coordinately and/or sequentially expressed during choriogenesis.

Since the libraries were prepared from random DNA fragments, independent isolates of a given gene will carry DNA sequences that extend for various distances away from the gene in both directions. This is clearly illustrated by the analysis of the β-globin clones (Figure 9). In some cases, the cloned DNA extends far enough in one direction to include a linked gene.

Although the possibility of linkage between different members of the rabbit α- or β-globin gene families has not been studied, evidence exists that such linkage occurs in other mammals, including mouse (Gilman and Smithies, 1968) and human (Clegg and Weatherall, 1976). For example, in humans the β-, δ- and γ-globin are genetically linked (Huisman et al., 1972; Clegg and Weatherall, 1976) on chromosome 11 (Deisseroth et al., 1978). In most human populations, the α-globin genes are present in two copies per haploid genome (Lehmann, 1970; Hollan et al., 1972) and are located on chromosome 16 (Deisseroth et al., 1977). Thus the close physical linkage between rabbit globin genes reported here is probably a general characteristic of mammalian globin genes. Clones bearing such genes can be used to study the precise organization of linked genes and the possible relation between linkage and control of gene expression. Moreover, a permanent library of clones bearing overlapping sequences will facilitate the isolation of the many linked genes that constitute a complex genetic locus.

Cloned segments of eucaryotic DNA can also be used to study the fine structure of genes. Most current methods of mammalian gene isolation involve partial purification of genomic DNA fragments generated by a limit restriction endonuclease digestion prior to cloning (Tilghman et al., 1977; Tonegawa et al., 1977). If the restriction endonuclease used to fragment genomic DNA cleaves within the coding sequence or noncoding intervening sequences of the gene of interest, the gene must be cloned in pieces. The rabbit α- and β-globin genes, which carry a single Eco RI site within the coding sequence (Maniatis et al., 1976; Salser et al., 1976; Liu et al., 1977), and the chicken ovalbumin gene, which carries at least one Eco RI site within each of three intervening sequences (Breathnach, Mandel and Chambon, 1977; Weinstock et al., 1977; Lai et al., 1978), illustrate this problem. If more than one Eco RI site is located within an intervening sequence, a portion of the chromosomal gene structure will remain unidentified. The procedure described here avoids these problems by cloning large pieces of randomly fragmented DNA that should carry intact genes including their intervening sequences.

Experimental Procedures

Materials

DNA polymerase I and T4 polynucleotide kinase were purchased from Boehringer Mannheim. T4 ligase was purchased from Bethesda Research Laboratories. Eco RI methylase prepared according to Greene et al. (1975) was provided by John Rosenberg. DNA polymerase from avian myeloblastosis virus (AMV reverse transcriptase) was provided by Dr. J. W. Beard and the Office of Program Resources and Logistics (Viral Cancer Program, NIH). Eco RI was prepared according to the procedure of Greene et al. (1975). Hae III and Alu I were prepared as described by Roberts et al. (1976) and Roberts (1976), respectively. Proteinase K was purchased from EM Labs. Pancreatic DNAase I was purchased from Worthington Biochemicals. NZ amine was purchased from Humko-Sheffield (Linnhurst, New Jersey). Nitrocellulose filters were purchased from Millipore. S-adenosyl-L-methionine was purchased from Sigma. α-^{32}P-deoxynucleoside triphosphates were purchased from New England Nuclear (~300 Ci/mmole) or ICN (120–200 Ci/mmole). Oligo$(T)_{12-18}$ was purchased from Collaborative Research.

Preparation of Bacteriophage λ DNA

Charon phage were grown essentially as described by F. R. Blattner in the detailed protocol that accompanies the Charon λ phages. Phage were purified as described in the above protocol and by Yamamoto et al. (1970).

For preparation of phage DNA, purified phage were dialyzed against 10 mM Tris-Cl (pH 8), 25 mM NaCl, 1 mM $MgSO_4$, brought to 0.2% SDS, 10 mM EDTA, heated to 65°C for 15 min and digested for 1 hr at 37°C with 50 μg/ml Proteinase K. The DNA was extracted several times with phenol, ether-extracted and dialyzed extensively against TSE [5 mM NaCl, 10 mM Tris-Cl (pH 8), 1 mM EDTA].

To prepare the end fragments of Charon 4 (see Figure 1), the cohesive ends were annealed by incubation for 1 hr in 0.1 M Tris-Cl (pH 8.0), 10 mM $MgCl_2$ at 42°C. Dithiothreitol (DTT) was added to 1 mM along with an excess of Eco RI and the reaction mix was incubated for 3 hr at 37°C. An aliquot was run on a 0.5% agarose gel to verify that digestion was complete. The DNA was extracted with phenol and then with ether. 50–70 μg of Eco RI-cleaved phage DNA were layered onto a 10–40% linear sucrose gradient [1 M NaCl, 20 mM Tris-Cl (pH 8.0), 10 mM EDTA] in a Beckman SW27 centrifuge tube. The gradient was centrifuged (at 27,000 rpm for 24 hr at 20°C) (Neal and Florini, 1972), and 0.5 ml fractions were collected using an ISCO ultraviolet flow cell. Fractions were analyzed on an agarose gel and those containing the 31 kb annealed end fragments were pooled.

To examine the restriction endonuclease cleavage patterns of DNAs from individual plaques, DNA was prepared as described above from phage grown in 4 ml cultures. Enough DNA was obtained to perform several restriction endonuclease digestions.

Drosophila DNA

Drosophila embryos (Canton S wild-type) aged 6–16 hr were collected, washed, frozen on dry ice and stored at −70°C. DNA was prepared according to Brutlag et al. (1977) with modifications. DNA from the CsCl gradient was dialyzed, digested with Proteinase K and phenol-extracted. The DNA was then brought to 5 M NaCl and chilled to 0°C to minimize the generation of molecules with single-stranded tails during shearing (Pyeritz, Schlegel and Thomas, 1972). The DNA was sheared by slowly

drawing it into a chilled 5 ml plastic syringe with a 20 gauge 1 in needle and expelling it as hard as possible into a 50 ml conical polyethylene tube on ice. The number of passes through the needle prior to the addition of S1 required to generate 20 kb fragments following the nuclease treatment varied from preparation to preparation. For this particular DNA preparation, three passes through a 20 gauge needle produced a mean size of 30 kb. The DNA was dialyzed against 0.5 M NaCl, 10 mM Tris-Cl (pH 8.0), 1 mM EDTA. Sodium acetate (pH 4.5) and $ZnSO_4$ were added to final concentrations of 50 mM and 2 mM, respectively. An amount of S1 nuclease sufficient to convert an equivalent amount of single-stranded λ DNA to small fragments as assayed by agarose gel electrophoresis was added. Following incubation for 1 hr at 37°C, the reaction mixture was extracted repeatedly with phenol and then with ether.

Silkmoth DNA

DNA was isolated from silkmoth Antheraea polyphemus pupae as previously described (Efstratiadis et al., 1976). Shearing and S1 nuclease digestion were performed as described above.

Rabbit Liver DNA

The liver of the New Zealand rabbit was removed and frozen in small pieces in liquid nitrogen. DNA was isolated using a modification of the Blin and Stafford (1976) procedure. After Proteinase K digestion, phenol extraction and dialysis, solid CsCl (0.95 g/ml) and ethidium bromide (1/10 vol of a 5 mg/ml solution) were added (final density 1.65 g/cm³). The solution was centrifuged (in a Ti60 rotor at 45,000 rpm for 60 hr at 20°C) to separate DNA from RNA and polysaccharides, and the DNA was collected as a viscous band by puncturing the side of the tube with a needle. Ethidium bromide was removed by several extractions with isopropanol equilibrated with saturated CsCl, followed by exhaustive dialysis against TSE. The molecular weight of the DNA was estimated by electrophoresis on neutral (Sharp, Sugden and Sambrook, 1973) and alkaline (McDonnel, Simon and Studier, 1977) 0.5% agarose gels, using bacteriophage λ Charon 4 DNA (46,200 bp, Blattner et al., 1977) and bacteriophage T4 DNA (171,000 bp, Kim and Davidson, 1974) as molecular weight standards. Both the duplex and single-stranded lengths of the rabbit DNA molecules were estimated to be >100,000 bp or nucleotides.

Partial endonuclease digestion conditions were established for the restriction enzymes Hae III and Alu I by performing a serial dilution of each enzyme in the presence of 1 μg of rabbit liver DNA in 1X restriction enzyme buffer [6 mM Tris-Cl (pH 7.5), 6 mM $MgCl_2$, 6 mM β-mercaptoethanol]. Reactions were incubated for 1 hr at 37°C, and the extent of digestion was estimated by electrophoresis on a 0.5% neutral agarose gel using Eco RI-digested Charon 4 DNA as a molecular weight standard. On the basis of this information, six large scale digests (330 μg DNA per reaction) were performed with 0.5, 1 and 2 times the estimated amount of enzyme yielding the maximum proportion of 20 kb fragments. The six digests were pooled, phenol-extracted and concentrated by ethanol precipitation.

Isolation of 20 kb Eucaryotic DNA

250-300 μg of sheared or enzymatically cleaved DNA in 0.5 ml of 10 mM Tris-Cl (pH 8.0), 10 mM EDTA were heated at 68°C for 20 min and sedimented through a 10-40% linear sucrose gradient as described above. Aliquots of fractions were analyzed by electrophoresis on a 0.5% agarose gel using Eco RI-digested Charon 4A DNA as a molecular weight standard. The fractions containing 19-20 kb DNA were pooled, dialyzed against TSE, concentrated by ethanol precipitation and resuspended in TSE.

Eco RI Methylation of Eucaryotic DNA

115 μg of 20 kb DNA were brought to a volume of 1 ml in 0.1 M Tris-Cl (pH 8.0), 10 mM EDTA, 6 μm S-adenosyl-L-methionine. Eco RI methylase (20 units in 1 ml) was added. Two 10 μl aliquots were taken before and after the addition of the enzyme and mixed with 0.5 μg of λ DNA. The eucaryotic DNA and the four 10 μl control reactions were incubated for 1 hr at 37°C. Each 10 μl aliquot containing λ DNA was mixed with 25 μl of a buffer containing 0.2 M Tris-Cl (pH 7.5), 0.1 M NaCl, 20 mM $MgCl_2$, and 2 mM DTT. Two of the aliquots (one each withdrawn before and after the addition of methylase) were mixed with Eco RI and incubated for 1 hr at 37°C. The other two corresponding aliquots were incubated without the addition of Eco RI (see Figure 3). The methylated DNA was phenol-extracted, ether-extracted, ethanol-precipitated, redissolved in 100 μl of 5 mM Tris-Cl (pH 7.5), dialyzed against the same buffer in a Schleicher and Schuell collodion bag, and evaporated to 40 μl under nitrogen.

Covalent Joining of Eco RI Linkers to Eucaryotic DNA

The synthesis of dodecamer linkers produces molecules with 5' hydroxyl ends (Scheller et al., 1977a). Since the ligase requires 5' phosphate ends, the first step in the joining reaction is to phosphorylate the linker. 5 μg of dodecamer linker in 10 μl of 66 mM Tris-Cl (pH 7.6), 10 mM $MgCl_2$, 1.0 mM ATP, 1.0 mM spermidine, 15 mM DTT, 200 μg/ml gelatin were added to 2 μl (10 units) of T4 kinase and incubated at 37°C for 1 hr. This reaction mixture was then added directly to 100 μl of eucaryotic DNA (~100 μg) in the same buffer. 5 μl (5 units) of T4 ligase were added and the reaction was incubated at room temperature for 6 hr. A 5 μl aliquot of the reaction mixture was electrophoresed on a 12% polyacrylamide Tris-borate-EDTA gel (Maniatis, Jeffrey and van de Sande, 1975) and the DNA was visualized by ethidium bromide staining. A successful ligation was evidenced by the presence of a series of linker oligomers, from dimers up to 14-mers.

The ligation reaction mixture was diluted to 500 μl with 10 mM EDTA, incubated for 15 min at 68°C, layered onto a 10-40% linear sucrose gradient and centrifuged as described above. The gradient fractions containing 20 kb DNA were identified by agarose gel electrophoresis, pooled, dialyzed against TSE and ethanol-precipitated. The DNA (25 μg) was resuspended in 100 μl of 5 mM NaCl and brought to 1X Eco RI buffer [0.1 M Tris-Cl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT], and 150 units of Eco RI were added. 3 μl of the reaction were mixed with 0.2 μg of λ DNA and both tubes were incubated for 1 hr at 37°C. The small scale reaction containing λ DNA was electrophoresed on a 0.5% agarose gel, in which a complete digest of the linkers attached to rabbit DNA was evidenced by a limit digestion pattern of λ DNA.

The Eco RI linkers were attached to 20 kb Drosophila and silkmoth DNAs using similar procedures, except that linker oligomers were removed on a Sepharose 2B column rather than on a sucrose gradient.

Ligation of Cohesive Ends

High cloning efficiencies were obtained using a 2 fold molar excess of Charon 4 arms to 20 kb eucaryotic DNA fragments. For example, in the case of the rabbit library, the reaction mixture contained 20.5 μg of eucaryotic DNA and 55 μg of purified Charon 4A arms in 300 μl of ligase buffer. The cohesive ends of the phage were annealed for a second time in $MgCl_2$, Tris and gelatin for 1 hr at 42°C before adding ATP, DTT, eucaryotic DNA and 19 μl (19 units) ligase. The mixture was incubated at 12°C for 12 hr. An aliquot was heated to 68°C, cooled and electrophoresed on a 0.3% agarose gel with Eco RI-digested Charon 4A DNA as a molecular weight standard. Successful ligation was evidenced by the absence of Charon 4A end fragments (~12 and 19 kb) and the presence of concatemeric DNA molecules larger than intact Charon 4A DNA.

In Vitro Packaging of Recombinant DNA into Phage Particles

In vitro packaging extracts were prepared as described by Sternberg et al. (1977), except that both types (A and B) of extracts which they describe were handled as 80 μl aliquots in 1.5 ml polypropylene tubes, which were frozen in liquid nitrogen and stored at -70°C. Proportionate amounts of lysozyme, then buffer B (Sternberg et al., 1977) and glycerol were added to each tube of

resuspended B extract cells and thoroughly stirred into the extremely viscous suspension using a glass micropipet.

Preparation and Testing of in Vitro Packaging Extracts for EK2 Experiments

NIH regulations require that in vitro packaging extracts "be irradiated with ultraviolet light to a dose of 40 phage lethal hits" before they can be used in EK2 level recombinant DNA experiments. Wavelengths between 250 and 280 nanometers are the most effective for photoinactivation (Hollaender, 1955). We use as an ultraviolet light source four 18 in 15 watt germicidal lamps (GE G15T8), which emit 1.3×10^6 erg/cm^2 – n (2200 μW/cm^2/sec) at a distance of 20 cm. The killing efficiency of this light source was calibrated as follows.

1 liter of NZCYM broth (1% NZ amine, 0.5% NaCl, 0.5% yeast extract, 0.1% casamino acids, 10 mM $MgSO_4$) was inoculated with 20 ml of an overnight culture of $\lambda C_I857Sam7$ (a heat-inducible λ lysogen). The culture was grown at 32°C to an OD_{550} of 0.3, transferred to a 42°C shaking water bath for 20 min to induce the lysogen and then incubated with shaking for 2.5 hr at 37°C. The cells were transferred to an enamel dish (32 cm × 18 cm × 5.5 cm) and irradiated at a distance of 20 cm with constant mixing on a rotary shaker platform. 1 ml aliquots were taken after 5, 10, 20, 30, 40 and 60 min of irradiation. The cells were lysed by the addition of two drops of chloroform and the DNA was removed by adding 10 μl of a 1 mg/ml DNAase I solution. The mixture was centrifuged for 2 min in a Brinkman Eppendorf centrifuge, and the number of surviving phage was determined by titering the supernatant on DP50SupF at 42°C under yellow light to avoid photoreactivation. Our yellow light source consists of a commercial "Gold" fluorescent lamp (GE) with a gold plexiglass filter. A plot of the log of survival versus time of irradiation extrapolates to 40 lethal hits at approximately 30 min of ultraviolet irradiation. Cells used in preparing EK2 extracts were ultraviolet-irradiated (as above) after heat induction and incubation at 37°C and prior to pelleting (see Sternberg et al., 1977).

Before any recombinant DNA was packaged, every preparation of extracts derived from ultraviolet-irradiated cells was tested for recombination and the presence of prophage. All of the steps described below were performed under yellow light. Extracts were tested for prophage excision and packaging by performing "mock" packaging reactions without exogenous DNA and plating on DP50SupF, which provides the appropriate suppressor for the S amber 7 mutation of the prophage (Sternberg et al., 1977). Plates were incubated at 42°C to inactivate the prophage repressor. No more than one tenth (30 μl) of one packaging reaction could be assayed on a single 10cm plate because the concentrated packaging mixture kills bacterial cells and thus masks the presence of prophage. According to NIH regulations, "the ratio of plaque-forming units without addition of exogenous λ DNA (endogenous virus) to plaque-forming units with exogenous DNA (exogenous virus) must be less than 10^{-6}." Exogenous viral DNA means recombinant DNA. We consistently find that this ratio is $< 10^{-6}$ for recombinant λ DNA and $< 10^{-8}$ using intact Charon 4A DNA in the assay. Without ultraviolet irradiation, the above ratio is $> 10^{-6}$.

NIH regulations further require that when the EK2 vector is packaged, "the ratio of am$^+$ phage (recombinants) to total phage must be less than 10^{-6}." We added Charon 4A DNA to the packaging extract and plated the packaged DNA on Su$^-$ and Su$^+$ strains. The ratio of pfu on Su$^-$ to pfu on Su$^+$ is a measure of the frequency of recombination with prophage DNA. We find that this ratio is $< 10^{-8}$; in fact, we have never observed a plaque resulting from recombination. On the basis of these and similar data obtained by N. Sternberg and L. Enquist (unpublished observations), the NIH has approved in vitro packaging for EK2 level experiments.

Packaging DNA for Libraries

To construct, for example, the rabbit library, 26 packaging reactions were performed, each containing 2.5 μg of recombinant DNA. After DNAase digestion and chloroform treatment, the packaged phage were purified and concentrated on a CsCl step gradient. The reactions were pooled, mixed with solid CsCl (0.5 g/ml), brought to a final volume of 30 ml with 0.5 g/ml CsCl in SM [0.1 M NaCl, 0.05 M Tris-Cl (pH 7.5), 10 mM $MgSO_4$, 0.01% gelatin], layered onto CsCl gradients (each one composed of 1.5 ml steps of 1.45, 1.5, 1.7 g/ml CsCl in SM, in an SW41 tube) and centrifuged (at 32,000 rpm for 1.5 hr at 4°C). 200 μl fractions were collected and phage were located by spotting dilutions of 10^{-3} to 10^{-4} onto a lawn of DP50SupF.

Library Amplification

The fractions containing phage were pooled and dialyzed against 0.1M NaCl, 50 mM Tris-Cl (pH 7.5), 10 mM $MgSO_4$. Gelatin was added to the dialysis bag to a concentration of 0.02% to stabilize the phage. In vitro packaged phage from the rabbit library were plated onto a fresh overnight of DP50SupF at a density of 10,000 phage per 15 cm diameter plate (75 plates total). The phage were preadsorbed with bacteria for 20 min at 37°C, mixed with 6.5 ml of NZCYM-DT (NZCYM medium with 0.01% diaminopimelic acid and 0.004% thymidine; Blattner et al., 1977), 0.5% top agar and plated. The plates were incubated for 14 hr at 37°C. (Plating was carried out in yellow light and the plates were wrapped in aluminum foil while growing to prevent photoreactivation.) The top agar was scraped into a sterile beaker and the plates were rinsed once with 3.75 ml of SM. The 285 ml lysate from 38 plates was mixed with 10 ml of chloroform and stirred slowly at room temperature for 20 min. The lysate was transferred to a 1 liter centrifuge bottle and centrifuged (in a Beckman J6 centrifuge at 52,000 rpm for 20 min at 4°C) to remove the top agar.

Screening the Libraries

Amplified libraries were screened using the in situ plaque hybridization technique of Benton and Davis (1977). 10,000 recombinant phage were plated on 3.1×10^8 exponential phase bacterial cells on 15 cm NZCYM petri dishes. To prevent top agar from adhering to the nitrocellulose filter when it was lifted from the plate (which tends to increase the background hybridization), plates were dried in a 37°C incubator for several hours or set on edge overnight to drain excess liquid. The use of 0.7% agarose rather than agar in the top agar layer also minimized this problem. The plates were incubated at 37°C for 14-16 hr, at which time the plaques were in contact but lysis was not confluent. Plates were refrigerated for an hour or longer before the filters were applied. Nitrocellulose filters were cut from rolls (HAMP 000 10) or circles (HAWP 142 50) of Millipore filter paper (type HA, pore size 0.45 μm) to fit easily over the agar plate. Phage and DNA were adsorbed to these filters in duplicate by placing two filters on each plate sequentially, 2 min for the first filter and 3 min for the second, at room temperature. The DNA was denatured and bound to the filters as described by Benton and Davis (1977).

To prepare the filters for hybridization to a labeled probe, they were wetted in about 10 ml per filter of 4X SET [1X SET = 0.15 M NaCl, 0.03 M Tris-Cl (pH 8), 2 mM EDTA] at room temperature for 30 min, washed for 3 hr at 68°C in about 10 ml per filter of 4X SET, 10X Denhardt's solution (10X Denhardt's solution = 0.2% bovine serum albumin, 0.2% polyvinylpyrolidone, 0.2% Ficoll, Denhardt, 1966) and 0.1% SDS, and prehybridized with continuous agitation for at least 1 hr at 68°C in about 4 ml per filter of 4X SET, 10X Denhardt's solution, 50 μg/ml denatured salmon sperm DNA, 10 μg/ml poly(A), 0.1% SDS and 0.1% sodium pyrophosphate. Denatured E. coli DNA (10–50 μg/ml) was included in the prehybridization mix when using labeled plasmid DNA as probe. The prehybridization and subsequent hybridization were carried out in a thermally sealed plastic bag. The filters were hybridized with agitation at 68°C in prehybridization solution containing ^{32}P-labeled hybridization probe at the concentrations and for the times indicated in the figure legends. After hybridization, the filters were washed once with agitation in about 15 ml per filter of 4X SET, 10X Denhardt's solution, 10 μg/ml poly(A), 0.1% SDS, 0.1% sodium pyrophosphate at 68°C for 1 hr; 3 times in a similar volume

of 3X SET, 0.1% SDS, 0.1% sodium pyrophosphate at 68°C for 15-30 min; twice in 1X SET, 0.1% SDS, 0.1% sodium pyrophosphate at 68°C for 15-30 min; and once in 4X SET at room temperature. In some cases, more stringent washing conditions (0.5X SET) were used as a final step. The filters were blotted dry, mounted on cardboard and exposed to preflashed Kodak XR5 X-ray film with Dupont Cronex;11R Xtra Life Lightning-plus intensifying screens at -70°C for 1-2 days.

Plaque Purification of Recombinant Phage

Plaques from the region of a plate corresponding to a positive on the autoradiogram were picked and suspended in 0.5 ml SM. The phage suspension was titered and the plate containing about 100 plaques was rescreened. The process of picking positives and rescreening was repeated until ~ 90% of the plaques on a plate gave positive signals after screening.

Plate lysates were prepared using 10^8 phage from an individual plaque on a 10 cm plate, and the lysates were harvested as described in Library Amplification. 10^{10}-10^{11} phage were usually recovered.

Hybridization Probes

cDNA plasmids were grown, purified and labeled in vitro by nick translation as previously described (Maniatis et al., 1976). Complementary DNA to globin and chorion mRNAs was synthesized as described by Efstratiadis et al. (1975) and Friedman and Rosbash (1977).

CsCl Sedimentation Equilibrium Analysis of Recombinant Phage

Phage from the amplified rabbit library (4×10^7 pfu) were mixed with 3×10^8 pfu of Charon 14 phage (Blattner et al., 1977) in a solution of CsCl, density = 1.514 g/ml, 10 mM Tris-Cl (pH 7.5) and 1.0 mM $MgSO_4$. The phage were banded by centrifugation to equilibrium in a Ti50 rotor, 5.38×10^8 g_{av} hr at 4°C. One drop fractions were collected from the bottom of the gradient into 0.5 ml SM for a total of 128 fractions. The fractions containing phage were titered on DP50SupF, a Su II + Su III strain, to determine the distribution of densities of phage in the library and also on Cla (SupO) to determine the position of the marker Charon 14, a nonamber phage. The position of a second marker, Charon 4A (representing the background phage in the library), was determined by counting the number of blue plaques formed on KH802, a Su II *lac*$^-$ strain, using Xgal plates. From the known sizes of Charon 14 and Charon 4A, a calibration curve was constructed relating fraction number, or buoyant density, to length of insert DNA. The average density of the recombinant phage provides a minimum estimate for the size of rabbit DNA inserts, since the empirical relationship between phage density and DNA molecular weight was established using λ DNA (50% G+C) (Davidson and Szybalski, 1971), while the base composition of rabbit DNA is 44.2% G+C (Kritskii, Arends and Nikolaev, 1967).

Recombinant DNA Safety

Experiments involving the cloning or propagation of bacteriophage λ carrying eucaryotic DNA were performed in accordance with the NIH Guidelines for recombinant DNA research. Drosophila and silkmoth cloning experiments were performed using EK1 vectors in P2 facilities located at the California Institute of Technology and Harvard University. Rabbit DNA cloning experiments were performed using the EK2 vector-host system λ Charon 4A/DP50SupF in a P3 facility at the California Institute of Technology. The relevant genotypes of EK2 vectors were verified prior to their use in recombinant DNA experiments.

Acknowledgments

The procedures for making and screening libraries were established concurrently in the laboratories of E. Davidson, N. Davidson, J. Bonner and L. Hood at Caltech. We benefited from the free exchange of information and ideas among these labs. We are grateful to D. Goldberg, B. Seed, D. Engel, R. Scheller and D. Anderson for suggestions and discussions, and to N. Davidson for critical readings of the manuscript. We thank F. Blattner for providing Charon phage strains, N. Sternberg and L. Enquist for providing in vitro packaging strains, C. K. Itakura and R. Scheller for providing Eco RI linkers, J. Rosenberg for providing Eco RI methylase, R. Robinson for providing labeled single-copy Drosophila DNA tracer, and A. Cortenbach for preparing media and materials. We thank W. Bender and D. Hogness for discussions and for communicating unpublished data on the Drosophila library characterization, and F. C. Kafatos for the use of facilities and his support. We thank J. Maniatis for help with the figures. This work was supported by an NSF grant to F. C. Kafatos and T. M.; an NIH grant to F. C. Kafatos; an American Cancer Society grant to T. M. and funds from an NIH Biomedical Research Support grant to the California Institute of Technology. R. C. H. was supported by a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. T. M. is the recipient of a Rita Allen Foundation career development award.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore by hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 16, 1978

References

Bahl, C. P., Marians, K. J., Wu, R., Stawinsky, J. and Narang, S. (1977). A general method for inserting specific DNA sequences into cloning vehicles. *Gene 1*, 81-92.

Bellet, A., Busse, H. and Baldwin, R. (1971). Tandem duplications in a derivative of phage lambda. In The Bacteriophage Lambda, A. J. Hershey, ed. (New York: Cold Spring Harbor Laboratory), pp. 501-513.

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. *Science 196*, 180-182.

Blattner, F. R., Williams, B. G., Blechl, A. E., Thompson, K. D., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, 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.

Blin, N. and Stafford, D. W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucl. Acids Res. 3*, 2303-2308.

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

Brutlag, D., Appels, R., Dennis, E. S. and Peacock, W. J. (1977). Highly repeated DNA in Drosophila melanogaster. *J. Mol. Biol. 112*, 31-47.

Carbon, J., Clarke, L., Ilgen, C. and Ratzkin, B. (1977). The construction and use of a hybrid plasmid gene bank in E. coli. In Recombinant Molecules: Impact on Science and Society, R. F. Beers and E. G. Bassatt, eds. (New York: Raven Press), pp. 335-378.

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. *Cell 9*, 91-99.

Clegg, J. B. and Weatherall, D. J. (1976). Molecular basis of thalassaemia. *Br. Med. Bull. 32*, 262-269.

Clissold, P. M., Arnstein, H. R. V. and Chesterton, C. J. (1977). Quantitation of globin mRNA levels during erythroid development in the rabbit and discovery of a new β-related species in immature erythroblasts. *Cell 11*, 353-361.

Davidson, N. and Szybalski, W. (1971). Physical and chemical characteristics of lambda DNA. In The Bacteriophage Lambda, A. D. Hershey, ed. (New York: Cold Spring Harbor Laboratory), pp. 45-82.

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., Lawrence, J., Creagen, R. and Kucherlapati, R. (1977). Localization of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. Cell *12*, 205-218.

Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. and Ruddle, F. (1978). Chromosomal localization of human β-globin gene on human chromosome 11 in somatic cell hybrids. Proc. Nat. Acad. Sci. USA *75*, 1456-1460.

Denhardt, D. T. (1966). A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. *23*, 641-646.

Efstratiadis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A. and Vournakis, J. N. (1975). Full length and discrete partial reverse transcripts of globin and chorion mRNAs. Cell *4*, 367-378.

Efstratiadis, A., Crain, W. R., Britten, R. J., Davidson, E. H. and Kafatos, F. (1976). DNA sequence organization in the lepidopteran Antheraea pernyi. Proc. Nat. Acad. Sci. USA *73*, 2289-2293.

Friedman, E. Y. and Rosbash, M. (1977). The synthesis of high yields of full-length reverse transcripts of globin mRNA. Nucl. Acids Res. *4*, 3455-3471.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. Cell *7*, 487-505.

Gilman, J. G. and Smithies, O. (1968). Fetal hemoglobin variants in mice. Science *160*, 885-886.

Glover, D. M., White, R. L., Finnegan, D. J. and Hogness, D. S. (1975). Characterization of six cloned DNAs from Drosophila melanogaster, including one that contains the genes for rRNA. Cell *5*, 149-157.

Greene, P. J., Poonian, M. S., Nussbaum, A. L., Tobias, L., Garfen, D. E., Boyer, H. W. and Goodman, H. M. (1975). Restriction and modification of a self complementary octanucleotide containing the EcoRI substrate. J. Mol. Biol. *99*, 237-261.

Heyneker, H. L., Shine, J., Goodman, H. J., Boyer, H., Rosenberg, J., Dickerson, R. E., Narang, D. A. and Riggs, A. D. (1976). Synthetic lac operator DNA is functional in vivo. Nature *263*, 748-752.

Hohn, B. and Murray, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Nat. Acad. Sci. USA *74*, 3259-3263.

Hollaender, A. (1955). Radiation Biology, *II* (New York: McGraw-Hill).

Hollan, S. R., Szelenyii, J. G., Brimhall, B., Duerst, M., Jones, R. T., Koler, R. D. and Stocklen, I. (1972). Multiple alpha chain loci for human haemoglobins: Hb J-Buda and Hb G-Pest. Nature *235*, 47-50.

Huisman, T. H. J., Wrightstone, R. M., Wilson, J. B., Schroeder, W. A. and Kendall, A. G. (1972). Hemoglobin Kenya, the product of a fusion of γ and β polypeptide chains. Arch. Biochem. Biophys. *153*, 850-853.

Jeffreys, A. J. and Flavell, R. A. (1977a). A physical map of the DNA regions flanking the rabbit β-globin gene. Cell *12*, 429-439.

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

Kafatos, F. C., Efstratiadis, A., Goldsmith, M. R., Jones, C. W., Maniatis, T., Reiger, J. C., Rodakis, G., Rosenthal, N., Sim, G. K., Thireos, G. and Villa-Komaroff, L. (1978). The developmentally regulated multigene families encoding chorion proteins in silkmoth. In Structure and Organization of Developmentally Regulated Genes, J. Schultz and F. Ahmad, eds. (New York: Academic Press), in press.

Kim, J. S. and Davidson, N. (1974). Electron microscope heteroduplex study of sequence relations of T2, T4 and T6 bacteriophage DNAs. Virology *57*, 93-111.

Kritskii, G., Arends, I. and Nikolaev, Y. (1967). Nucleotide composition of DNA fractions of bone marrow in normal and X-irradiated animals. Biochemistry (USSR) *32*, 372-376.

Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Catteral, J. F. and O'Malley, B. W. (1978). The ovalbumin gene: structural sequences in native chick DNA are not contiguous. Proc. Nat. Acad. Sci. USA *75*, 2205-2209.

Leder, P., Tiemeier, D. and Enquist, L. (1977). EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. Science *196*, 175-177.

Lehmann, H. (1970). Different types of alpha thalassemia and significance of haemoglobin Bart's in neonates. Lancet *2*, 78-80.

Liu, A. Y., Paddock, G. V., Heindell, H. C. and Salser, W. (1977). Nucleotide sequences from a rabbit alpha globin gene inserted in a chimeric plasmid. Science *196*, 192-197.

Mandel, M. and Higa, A. (1970). Calcium-dependent bacteriophage DNA interaction. J. Mol. Biol. *53*, 159-162.

Maniatis, T., Jeffrey, A. and van de Sande, H. (1975). Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry *14*, 3787-3794.

Maniatis, T., Sim, G. K., Efstratiadis, A. and Kafatos, F. C. (1976). Amplification and characterization of a β-globin gene synthesized in vitro. Cell *8*, 163-182.

Maniatis, T., Sim, G. K., Kafatos, F. C., Villa-Komaroff, L. and Efstratiadis, A. (1977) An approach to the study of developmentally regulated genes. In Molecular Cloning of Recombinant DNA (New York. Academic Press), pp. 173-203.

McDonnel, M., Simon, M. N. and Studier, W. F. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. *110*, 119-146.

Melderis, H., Steinheider, G. and Ostertag, W. (1974). Evidence for a unique kind of α type globin chain in early mammalian embryos. Nature *250*, 774-776.

Neal, M. W. and Florini, J. R. (1972). Effect of sucrose gradient composition on resolution of RNA species. Anal. Biochem. *45*, 271-276.

Pyeritz, R., Schlegel, R. and Thomas, C., Jr. (1972). Hydrodynamic shear breakage of DNA may produce single-chained terminals. Biochim. Biophys. Acta *272*, 504-509.

Roberts, R. J. (1976). Restriction endonucleases. CRC Crit. Rev. Biochem. *4*, 123-164.

Roberts, R. J., Myers, P. A., Morrison, A. and Murray, K. (1976). A specific endonuclease from Arthobacter luteus. J. Mol. Biol. *102*, 157-165.

Rudkin, G. T. (1972). Replication in polytene chromosomes. In Results and Problems in Cell Differentiation, *4*, W. Beerman, ed. (New York: Springer-Verlag), pp. 59-85.

Salser, W., Bowen, S., Brown, D., Adli, F. E., Federoff, N., Fry, K., Heindell, H., Paddock, G., Poor, R., Wallace, B. and Whitcome, P. (1976). Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. *35*, 23-35.

Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. and Itakura, K. (1977a). Chemical synthesis of restriction enzyme recognition sites useful for cloning. Science *196*, 177-180.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977b). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic eco RI sites. Science *196*, 197-200.

Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. and Goodman, H. M. (1977). Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. Nature *270*, 486-494.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973). Detection of two restriction endonuclease activities in Haemophilus parain-

fluenzae using analytical agarose ethidium bromide electrophoresis. Biochemistry *12*, 3055–3063.

Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. and Goodman, H. M. (1977). Construction and analysis of recombinant DNA for human chorionic somatomammotropin. Nature *270*, 494–499.

Sim, G. K., Efstratiadis, A., Jones, C. W., Kafatos, F. C., Koehler, M., Kronenberg, H. M., Maniatis, T., Regier, J. C., Roberts, B. F. and Rosenthal, N. (1978). Studies on the structure of genes expressed during development. Cold Spring Harbor Symp. Quant. Biol. *42*, 933–945.

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

Steinheider, G., Melderis, H. and Ostertag, W. (1975). Embryonic ε chains of mice and rabbits. Nature *257*, 714–716.

Steinheider, G., Melderis, H. and Ostertag, W. (1977). In International Symposium on the Synthesis, Structure and Function of Hemoglobin, H. Martin and L. Novicke, eds. (Munich: Lehmans), pp. 222–235.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). In vitro packaging of a λ *Dam* vector containing Eco RI DNA fragments of E. and phage P1. Gene *1*, 255–280.

Thomas, M., Cameron, J. R. and Davis, R. W. (1974). Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. Proc. Nat. Acad. Sci. USA *71*, 4579–4583.

Tilghman, S. M., Tiemeier, D. C., Polsky, F., Edgell, M. H., Seidman, J. G., Leder, A., Enquist, L. W., Norman, B. and Leder, P. (1977). Cloning specific segments of the mammalian genome: bacteriophage λ containing mouse globin and surrounding gene sequences. Proc. Nat. Acad. Sci. USA *74*, 4406–4410.

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. and Leder, P. (1978). Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc. Nat. Acad. Sci. USA *75*, 725–729.

Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977). Cloning of an immunoglobulin variable region gene from mouse embryo. Proc. Nat. Acad. Sci. USA *74*, 3518–3522.

Ullrich, A., Shine, J., Chirgwin, J., Dictet, R., Tischer, E., Rolfes, W. J. and Goodman, H. J. (1977). Rat insulin genes: construction of plasmids containing the coding sequences. Science *196*, 1313–1319.

Weigle, J., Meselson, M. and Paigen, K. (1959). Density alterations associated with transducing ability in bacteriophage lambda. J. Mol. Biol. *1*, 379–386.

Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel, R. (1977). Intragenic spacers interrupt the ovalbumin gene. Proc. Nat. Acad. Sci. USA *75*, 1299–1303.

Wensink, P. C., Finnegan, D. J., Donelson, J. E. and Hogness, D. S. (1974). A system for mapping DNA sequences in the chromosomes of Drosophila melanogaster. Cell *3*, 315–325.

Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L. and Treiber, C. (1970). Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. Virology *40*, 734–744.

Young, M. and Hogness, D. (1977). A new approach for identifying and mapping structural genes in Drosophila melanogaster. In Eucaroytic Genetic Systems, ICN-UCLA Symposia on Molecular and Cellular Biology, *VIII* (New York: Academic Press), pp. 315–331.

Note Added in Proof

The NIH has recently revised the rules for in vitro packaging in EK2 systems using phage λ vectors. In particular, ultraviolet irradiation of extracts to a dose of 40 phage-lethal hits and assaying extracts without the addition of exogenous λ DNA are no longer required (for current rules, consult Office of Recombinant DNA Activities).

Chapter 2

Cell, Vol. 18, 1273–1283, December 1979, Copyright © 1979 by MIT

The Linkage Arrangement of Four Rabbit β-Like Globin Genes

Elizabeth Lacy, Ross C. Hardison, Diana Quon and Tom Maniatis
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

Four different regions of rabbit β-like globin gene sequences designated β1, β2, β3 and β4 were identified in a set of clones isolated from a bacteriophage λ library of chromosomal DNA fragments (Maniatis et al., 1978). Restriction mapping and blot hybridization (Southern, 1975) studies indicate that a subset of these clones containing β1 and β2 hybridizes to an adult β-globin cDNA clone (Maniatis et al., 1976) more efficiently than to a human γ-globin cDNA clone (Wilson et al., 1978), while another subset containing β3 and β4 displays the converse hybridization specificity. β1 was identified as the adult β-globin gene, while β2, β3 and β4 have not been identified with any known rabbit globin polypeptides. Cross-hybridization and transcriptional orientation experiments indicate that the set of β-like gene clones contains overlapping restriction fragments encompassing 44 kb of rabbit chromosomal DNA. In addition, all four genes have the same transcriptional orientation and are arranged in the order 5′-β4-β3-β2-β1-3′.

Introduction

The genes encoding the mammalian α-like and β-like polypeptide chains of hemoglobins comprise a relatively small gene family whose members are differentially expressed during development (Kitchen and Brett, 1974; Bunn, Forget and Ranney, 1977). Genetic studies indicate that the human α-like and β-like globin genes are on different chromosomes but that different β-like genes are closely linked (Weatherall and Clegg, 1972; Bunn et al., 1977). Similar conclusions have been reached with the mouse globin system (Gilman and Smithies, 1968; Russell and MacFarland, 1974). Direct physical linkage between different human β-like globin genes was recently demonstrated by genomic blotting experiments (Flavell et al., 1978; Mears et al., 1978; Fritsch, Lawn and Maniatis, 1979; Little et al., 1979; Tuan et al., 1979) and by analysis of clones containing the genes (Lawn et al., 1978; Fritsch et al., 1979). In this paper we describe the arrangement of four different β-like genes within a 44 kb segment of rabbit chromosomal DNA. Analysis of overlapping recombinant phage has made it possible to establish the linkage arrangement, intergene distance and transcriptional orientation of the genes. In the accompanying manuscript we describe the use of embryonic and adult hybridization probes and detailed structural studies as a means to identify each of the four genes (Hardison et al., 1979).

Results

Isolation of Phage Recombinants Containing Rabbit β-Like Globin Gene Sequences

We previously described the construction and characterization of a bacteriophage λ library of random, high molecular weight fragments of rabbit liver DNA (Maniatis et al., 1978). By screening this library with ^{32}P-labeled cDNA prepared from adult rabbit reticulocyte mRNA, we obtained nine different phage recombinants containing β-like globin gene sequences. Four of these recombinants displayed intense signals after hybridization with a cloned adult rabbit β-globin cDNA probe (pRcβ1; Maniatis et al., 1976) while the remaining five clones displayed a relatively weak signal. The intensity of the signal for the two sets of clones displayed the converse specificity, however, when they were hybridized to a cloned human γ-globin cDNA probe (pHcγ-151; Wilson et al., 1978) (data not shown). We will refer to the four clones which efficiently hybridize the adult rabbit β-globin probe as "adult" and designate the clones λRβGn (see Table 1 for an explanation of this nomenclature). The five clones which hybridize most efficiently to the human γ-globin probe will be referred to as "non-adult" and will be designated λRβ′Gn.

Physical Linkage between the Adult (β) and Non-adult (β′) Globin Gene Sequences

The arrangement of β-like sequences in the cloned DNA was determined by constructing detailed restriction maps and determining which fragments hybridized to the β-globin probes using the blot hybridization procedure (Southern, 1975). In addition to using standard single and double restriction enzyme digestion procedures to derive the restriction maps, we used a variety of other methods which are discussed in Experimental Procedures. A restriction map derived for the sequences present in three of the adult clones is shown in Figure 1. We have designated the regions of these clones which hybridize to β-globin sequence probes as β1, β2 and β3. β1 is cleaved by Eco RI into fragments of 2.6 and 0.8 kb. Since these fragments are, respectively, the sizes of the 5′ and 3′ adult β-globin Eco RI fragments found in genomic DNA (Jeffreys and Flavell, 1977a, 1977b), β1 is the adult β-globin gene. A characterization of β2 and β3 will be presented in the accompanying paper (Hardison et al., 1979). The restriction map in Figure 1 reveals that β2 and β3 lie 5′ to β1, and that β2 is separated from β1 by approximately 7.4 kb and from β3 by approximately 5 kb of DNA.

Figure 2 shows blot hybridization data used to lo-

Cell
1274

cate the globin sequences in the non-adult (β') recombinant phage DNAs. Each β' clone contains two Eco RI fragments which hybridize to the human γ-globin probe, in addition to a number of nonhybridizing fragments which are common to all five clones (Figures 2A and 2B). Although the sizes of the hybridizing fragments differ in each clone, only two of these (2.8 and 3.3 kb) are observed in whole genome blotting experiments (Figure 2C, lane 2), indicating that only one clone ($\lambda R\beta'G8$; Figure 2B, lane 4) contains both of the Eco RI fragments present in genomic DNA. The other hybridizing bands observed in the genomic blots of Figure 2C, lane 2, correspond to the 2.6 kb 5′ fragment of gene $\beta 1$ (lane 1), the 6.3 fragment of $\beta 2$ and to three other fragments which are not represented in our collection of clones and may therefore correspond to other β-like globin genes (for discussion see Hardison et al., 1979). The other, smaller hybridizing fragments observed in the β' clones (Figure 2B) can be explained in the following way. When the rabbit DNA library was made, Eco RI linkers were attached to the ends of 20 kb fragments produced by nonlimit restriction endonuclease cleavage of high molecular weight genomic DNA. If the random enzyme cleavage occurred between two Eco RI sites in a region containing β-globin sequence, linker attachment will produce a β-globin-hybridizing Eco RI fragment in the cloned genomic DNA which is shorter than the one present in total rabbit genomic DNA. The shorter Eco RI fragments are therefore the terminal Eco RI fragments of the DNA inserts. To locate the relative positions of the β-like globin gene sequences precisely, a restriction map was derived for $\lambda R\beta'G8$ (Figure 3). The map shows that both of the regions which hybridize to the human γ-globin probe span approximately 1.5 kb of DNA and are separated by approximately 8 kb of nonglobin DNA.

Comparison of the restriction maps of the adult and non-adult β-like globin gene clones strongly suggested that the two sets of recombinant phage contain overlapping segments of rabbit chromosomal DNA. Specifically, Eco RI fragments of 0.36 and 0.7 kb and a Bam HI fragment of 2.9 kb are found in the $\beta 3$ region of $\lambda R\beta G5$ (Figure 1) and near one of the β-like globin sequences in $\lambda R\beta'G8$ (Figure 3). Cross-hybridization experiments between DNA from the adult and non-adult clones were carried out to demonstrate unambiguously that there is overlap between the two sets of clones. The Eco RI digestion patterns of three different clones ($\lambda R\beta G5$, $\lambda R\beta'G9$, $\lambda R\beta'G8$) are shown in Figure 4A. Fragments of 0.7 and 0.36 kb are observed in all three DNAs. In addition, $\lambda R\beta G5$ and $\lambda R\beta'G9$ (Figure 4A, lanes 1 and 2, respectively) both contain fragments of 2.06 and 2.13 kb while $\lambda R\beta'G8$ contains a 2.0 kb fragment. The Eco RI-digested DNA of Figure 4A was blotted onto nitrocellulose and the filters were probed with in vitro labeled DNA from $\lambda R\beta G5$ (Figure 4B) or $\lambda R\beta'G8$ (Figure 4C). Labeled

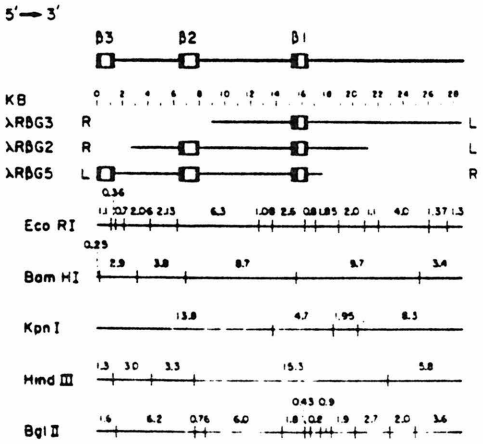

Figure 1. Restriction Endonuclease Cleavage Maps of the Clones $\lambda R\beta G3$, $\lambda R\beta G2$ and $\lambda R\beta G5$

The rabbit DNA sequences contained in these three clones overlap, and together they comprise a continuous segment of chromosomal DNA approximately 29 kb in length as illustrated in the top line. The regions which hybridize to β-globin sequences are designated $\beta 1$, $\beta 2$ and $\beta 3$. The black boxes are mRNA-coding regions and the white boxes are noncoding intervening sequences, as determined in the accompanying paper (Hardison et al., 1979). The map is drawn so that the 5′–3′ orientation determined in Figures 7 and 8 is from left to right. The line diagrams beneath the schematic map of the entire 29 kb of DNA represent the amount of rabbit genomic DNA carried in each clone. (L and R) indicate the left and right arms, respectively, of the λCh4A vector. (See Figure 4 for restriction maps of the vector DNA.) A map is not presented for each clone individually, but for the composite rabbit DNA insert. The size of each fragment is given in kb.

$\lambda R\beta G5$ DNA hybridizes intensely to fragments of 2.8, 0.7 and 0.36 kb in $\lambda R\beta'G9$ and $\lambda R\beta'G8$ DNA as well as to bands in the 2 kb regions in both DNAs (Figure 4B, lanes 2 and 3). Thus these fragments must contain sequences common to all three clones. Conversely, labeled $\lambda R\beta'G8$ DNA hybridizes to fragments of 0.7, 0.36 and 2.06 and/or 2.13 kb in $\lambda R\beta G5$ plus another fragment of 1.1 kb (Figure 4C). The results of this experiment, in conjunction with the mapping data of Figures 1 and 3, establish the linkage map shown in Figure 5. Fragments of 2.13, 2.06 (2.0 in $\lambda R\beta'G8$), 0.7 and 0.36 kb are present in the region of overlap and the $\beta 3$ gene sequence is present in both $\lambda R\beta G5$ and all of the β' clones. Other hybridizing bands observed in Figure 4 are a result of either globin cross-homology or hybridization between repeated elements within the gene cluster. A detailed characterization of these repeated elements will be reported elsewhere (J. Shen and T. Maniatis, manuscript in preparation). Thus a total of 44 kb of rabbit genomic DNA is present in the set of nine overlapping recombinant phage. This region of the chromosome contains four globin sequences, each separated by 5–8 kb of

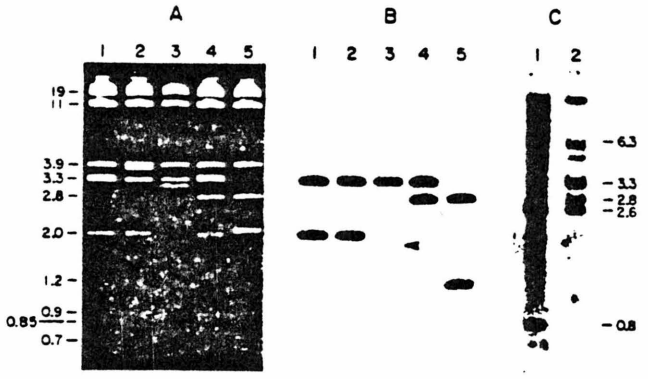

Figure 2. Blot Hybridization Study of the Non-adult β-Like Globin Gene Clones

(A) Fluorograph of ethidium bromide-stained gel. DNAs (0.3 μg each) from clones λRβ'G2 (lane 1), λRβ'G3 (lane 2), λRβ'G7 (lane 3), λRβ'G8 (lane 4) and λRβ'G9 (lane 5) were digested with Eco RI and electrophoresed in a 1% agarose gel calibrated with markers of bacteriophage λ (strain cI ts857 S7; Pirrotta et al., 1971) cut with Hind III (for sizes see Robinson and Landy, 1977) and φX174RFI DNA cut with Hae III (Sanger et al., 1977). The gel was stained with ethidium bromide and visualized by fluorescence during ultraviolet irradiation. Sizes determined for the Eco RI fragments are given in kb. Clones λRβ'G8 and λRβ'G9 also contain a faint 0.36 kb fragment not visible in this photograph. They can be seen in Figure 5.

(B) Autoradiogram of the DNA in the gel of (A) hybridized with the human γ-globin cDNA probe. DNA from the agarose gel in (A) was transferred to nitrocellulose (Southern, 1975) and hybridized with in vitro labeled pHcγ-151 (Wilson et al., 1978), a human γ-globin cDNA plasmid. The faintly hybridizing 1.8 kb Eco RI fragment from λRβ'G7 is indicated by the arrow.

(C) Autoradiogram of total rabbit genomic DNA hybridized with the rabbit β- and human γ-globin cDNA probes. Rabbit fetal DNA (15 μg per lane) was digested with Eco RI, electrophoresed on a 1% agarose gel, transferred to nitrocellulose and hybridized with a ^{32}P-labeled rabbit adult β-globin cDNA plasmid pRcβ1 (lane 1) or human γ-globin plasmid pHcγ-151 (lane 2).

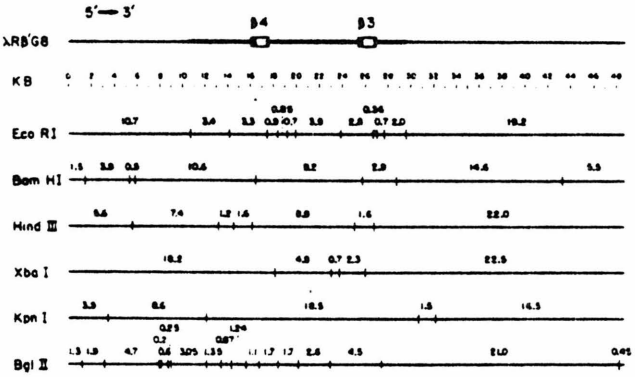

Figure 3. Restriction Endonuclease Cleavage Map of Clone λRβ'G8

The map is drawn so that the 5' → 3' orientation determined in Figure 8 is from left to right. Consequently, the λCh4A vector short arm (10.7 kb Eco RI fragment) is on the left and the long arm (19.2 kb Eco RI fragment) is on the right, which is the opposite of the usual convention for drawing maps of λ DNA (Blattner et al., 1977). The rabbit DNA insert is denoted by a thickened line. The black boxes are mRNA-coding regions of the β-like globin genes and the white boxes are noncoding intervening sequences as determined in the accompanying paper (Hardison et al., 1979). Sizes of the fragments are in kb.

DNA. As shown in Figure 5, we have designated these four sequences as β1, β2, β3 and β4.

Transcriptional Orientation of the Linked β-Like Globin Sequences

One question raised by the linkage arrangement established above is whether β2, β3 and β4 have the same transcriptional orientation as β1, the adult β-globin gene. We determined the transcriptional orientation of β2 by using hybridization probes specific for the 5' and 3' sequences in the adult β-globin mRNA. These probes were derived from the adult β-globin cDNA plasmid pRcβ1. Because of uncertainties in the relative locations of restriction sites within the β-globin cDNA and the β2 genomic sequence, it was desirable to prepare two types of hybridization probes. The regions of the β-globin mRNA sequence covered by the two types of probes are indicated in Figure 6A. Blots of λRβG2 DNA digested with Hind III and Eco RI were prepared and hybridized with the probes described in Figure 6. A map of the Hind III and Eco RI sites in λRβG2 DNA is shown in Figure 6B. Hybridization with the complete β-globin cDNA plasmid reveals four bands corresponding to fragment sizes of 5.0, 2.6, 1.33 and 0.8 kb (Figure 6C, lane 6). The 2.6 and 0.8 kb fragments contain the 5' and 3' adult β-globin mRNA sequences, respectively (Jeffreys and Flavell, 1977a). As expected, only the 2.6 kb fragment is detected by both types of 5'-specific probes (Figure 6C, lanes 1 and 4). In addition, the 0.8 kb fragment hybridizes preferentially to both 3'-specific probes. Even though unlabeled 5'-specific competitor DNA is included in the hybridization reactions with the 3' probes, significant labeling of the 2.6 kb fragment is observed. This is a result of the difficulty in obtaining homogeneous hybridization probes by gel purification.

As shown in Figure 6B, gene β2 is cleaved into

Cell
1276

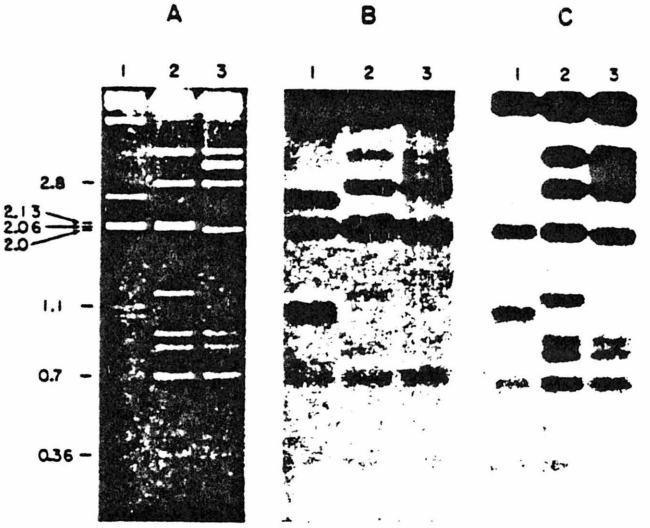

Figure 4. Blot Hybridization Data Showing Overlap between the Rabbit DNA in the β-Like Globin Gene Clones

(A) Fluorograph of ethidium bromide-stained agarose gel. DNA (0.7 μg each) from clones λRβG5 (lane 1), λRβ′G9 (lane 2) and λRβ′G8 (lane 3) was digested with Eco RI and fractionated on a 1.4% agarose gel.

(B) Autoradiogram showing hybridization to λRβG5. A gel identical to that in (A) was blotted onto nitrocellulose and probed with labeled DNA from clone λRβG5.

(C) Autoradiogram showing hybridization to λRβ′G8. The gel in (A) was blotted onto nitrocellulose and hybridized with labeled DNA from clone λRβ′G8. Fragments containing homologous sequences in the two sets of clones are denoted in the margin by their size in kb.

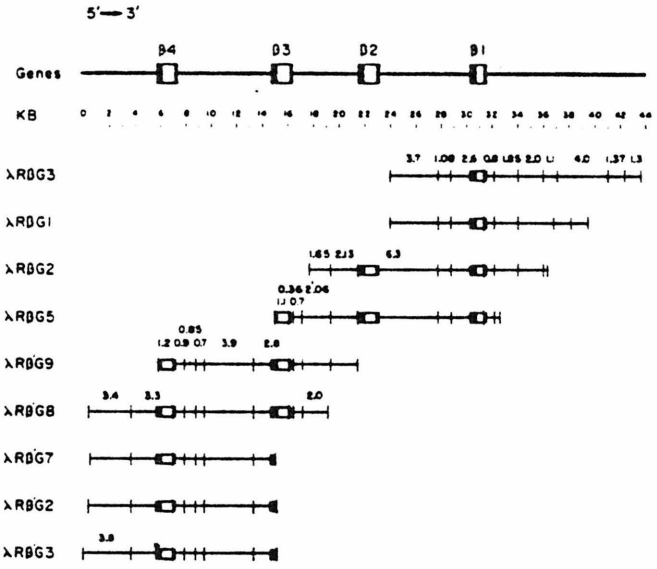

Figure 5. Linkage Arrangement of Nine Overlapping Clones Containing Four Linked β-Like Globin Genes

The nine isolated clones containing β-like globin genes are arranged to show their overlapping Eco RI fragments. Sizes of the fragments are given in kb; only the rabbit DNA inserts from the clones are shown. The boxes designate the positions of the globin genes, with the black portions containing coding sequences and the white containing intervening sequences (Hardison et al., 1979).

fragments of 1.33 and 5.0 kb. Both 5′-specific probes hybridize exclusively with the 1.33 kb fragment. The 3′ Eco RI probe preferentially detects the 5.0 kb relative to the 1.33 kb fragment (Figure 6C, lane 2). Thus, according to Figure 6B, β2 and β1 have the same transcriptional orientation. When the 3′-specific probe which consists primarily of 3′ untranslated β-globin mRNA sequences (3′ Hae III, Figure 6A) was used, the β2 5.0 kb fragment was not detected but the β1 0.8 kb fragment hybridized strongly (Figure 6C, lane 5). This finding is not unexpected since other studies have shown that the 3′ untranslated region of β-like mRNA sequences is the most divergent among different globin genes (Kafatos et al., 1977; Forget et al., 1979).

A blot hybridization analysis similar to that shown in Figure 6 was used to establish the transcriptional orientation of β3 and β4. λRβ′G8 DNA was digested with Bam HI and Hind III and hybridized with the 5′- and 3′-specific probes derived from the human γ-

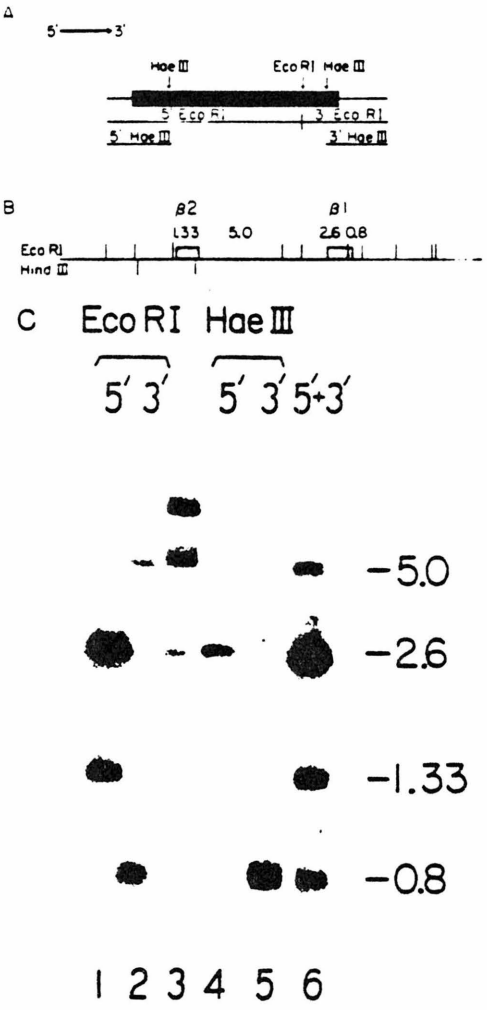

Figure 6. Transcriptional Orientation of β2

(A) A schematic diagram of the adult rabbit β-globin cDNA insert in the plasmid pRcβ1 showing the regions of mRNA sequence contained in the two types of 5'- and 3'-specific hybridization probes generated by Eco RI or Hae III cleavage. The protein-coding sequences (thick horizontal line) and the untranslated mRNA sequences (thin lines) are indicated. The pMB9 vector sequences are not shown. The 5' and 3' Eco RI probes were prepared by cleavage of pRcβ1 DNA with Hha I and Eco RI to generate a 1.05 kb fragment containing 361 nucleotides (nt) of 5' coding and 43 nt of 5' noncoding β-globin message sequences, and a 0.75 kb fragment containing 77 nt of 3' coding and 95 nt of 3' noncoding message sequence. The Hae III probes were generated by digesting pRcβ1 DNA with Hae III to produce a 0.57 kb fragment containing 79 nt of the extreme 5' coding sequence and 43 nt of 5' noncoding sequence, and a 0.79 kb fragment containing 26 nt of 3' coding and 95 nt of 3' noncoding sequence (Maniatis et al., 1976; Efstratiadis, Kafatos and Maniatis, 1977).

(B) A map of λRβG2 showing the location of restriction sites used in the analysis of the transcriptional orientation of β2. Only the rabbit globin cDNA plasmid pHcγ-151, as described in the legend to Figure 7. A map of the Bam HI and the Hind III sites surrounding β3 and β4 is presented for reference in Figure 7A. When the blots are hybridized with the probe containing both 5' and 3' mRNA sequences, bands corresponding to fragments of 0.3, 0.6, 1.0, 1.6 and 8.4 kb are observed (Figure 7B). The 1.6 kb band is very faint, indicating substantial divergence in this region between β4 and γ. When the 5'-specific probe is used, the 0.6, 0.3 and 1.6 kb fragments hybridize. (Although the 1.6 kb band is not visible in this reproduction, the band is clearly present on the original autoradiogram.) Conversely, the 8.4 and 1.0 kb fragments hybridize more efficiently to the 3'-specific probe than to the 5' probe. Thus β3 and β4 are transcribed from the same DNA strand. Since β3 is present in both λRβ'G8 and λRβG5 and since the latter clone contains both β2 and β1 sequences, we can conclude from the data of Figures 6 and 7 and the map in Figure 5 that all four genes are transcribed from the same DNA strand in the order 5'-β4-β3-β2-β1-3'.

Correspondence between Cloned and Total Genomic DNA Containing β-Like Globin Gene Sequences

Although maps of the human β-like globin gene cluster derived from genomic blotting experiments (Flavell et al., 1978; Mears et al., 1978; Fritsch et al., 1979; Little et al., 1979) and from cloned DNA (Lawn et al., 1978; Fritsch et al., 1979) are indistinguishable, rearrangements have been observed in some Charon 4A clones containing the human α (J. Lauer, personal communication) and γ (E. Fritsch, personal communication) globin genes. To show that the arrangement of sequences in cloned and total rabbit genomic DNA is the same, we compared the sizes of the restriction fragments in total genomic DNA to those predicted by the map of the gene cluster shown in Figure 9. By choosing the appropriate hybridization probes and restriction digests, we were able to analyze the entire gene cluster. Digests of a mixture of cloned DNAs from λRβG3, λRβG2 and λRβ'G8, which includes the entire isolated gene cluster, were blotted in parallel with the equivalent digests of total chromosomal DNA.

DNA insert is shown. The shadowed boxes cover the fragments containing globin mRNA-homologous sequences.

(C) An autoradiogram showing the results of the β2 transcriptional orientation experiment. 3.5 μg of λRβG2 DNA were digested with Eco RI and Hind III, loaded onto a single 7.5 cm wide slot of a 1.4% agarose gel, electrophoresed and transferred to nitrocellulose. The blot was cut into 0.5 cm strips, each of which was hybridized to one of six probes: (lane 1) Eco RI 5'; (lane 2) Eco RI 3'; (lane 3) λRβG2; (lane 4) Hae III 5'; (lane 5) Hae III 3'; (lane 6) total pRcβ1. All the DNAs used as probes were labeled in vitro with ^{32}P by nick translation. To compete out hybridization due to contaminating 3' or 5' sequences, blots hybridized to 5' probes were prehybridized with the corresponding unlabeled 3' probe and blots hybridized to 3' probes were prehybridized with the corresponding unlabeled 5' probe.

Cell
1278

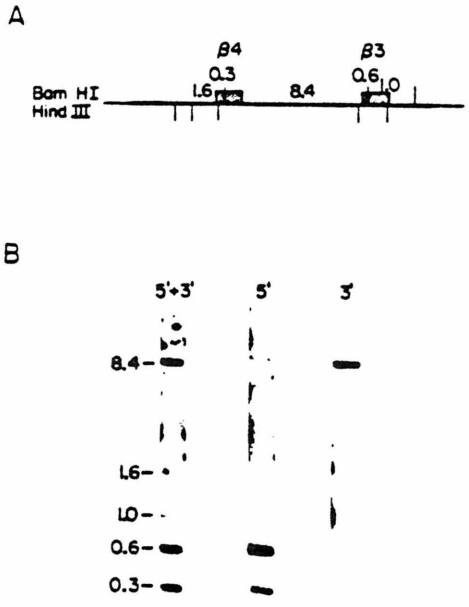

Figure 7. Transcriptional Orientation of Genes β3 and β4

(A) A map of λRβ'G8 showing the location of restriction sites used in the transcriptional orientation analysis. Only the rabbit DNA insert is shown. The shadowed boxes cover the fragments containing globin mRNA-homologous sequences.

(B) Autoradiogram showing the results of the β3 and β4 transcriptional orientation experiment. DNA (0.5 μg) from clone λRβ'G8 was digested with Bam HI plus Hind III, electrophoresed in triplicate on a 1% agarose gel calibrated with pBR322 plasmid DNA restriction fragments and transferred to nitrocellulose. The lanes were hybridized with probes for either the entire 5' + 3', the 5' end or the 3' end of the γ-globin cDNA clone. The orientation of the double-stranded cDNA insert in pHcγ-151 is 5' → 3' clockwise (B. Forget, personal communication) in the usual map orientation of pMB9 (Rodriguez et al., 1976). Cleavage with Bam HI therefore generates an approximately 6 kb fragment containing the 5' portion of γ-globin cDNA (through codons 98–100) and an approximately 0.6 kb fragment containing the 3' portion (codons 98–100 through the protein-coding region). These fragments, as well as the entire cDNA plasmid, were labeled in vitro with ^{32}P and used as probes in this experiment.

In all cases, the fragments which hybridize in total genomic DNA are identical in size to those in the cloned DNA (data not shown). These data indicate clearly that no detectable rearrangements occurred within the rabbit β-related globin gene cluster during preparation, isolation or propagation of the clones examined. Because of the limited sensitivity of this assay, however, we cannot rule out the possibility that small rearrangements occurred during cloning.

Although no rearrangements occurred in the three clones described above, we have detected a deletion near the β1 gene in λRβG1 DNA. The location of this deletion is revealed by comparing the maps of λRβG3 and λRβG1 which contain the same region of rabbit chromosomal DNA (Figure 8). The fact that the terminal Eco RI fragments from the rabbit DNA inserts in the two clones are identical in size suggests that λRβG1 is a deletion product of λRβG3. A genomic blot of the rabbit DNA used in constructing the library was probed with the cloned β1 intervening sequence. Only the 9.7 kb Bam HI fragment characteristic of λRβG3 DNA was observed, indicating that the deletion in λRβG1 DNA must be an artifact of cloning. It is interesting to note that a sequence present in the 2.0 kb Eco RI fragment of λRβG3 is also found in the 4.0 kb fragment (J. Shen and T. Maniatis, manuscript in preparation). If this is a direct repeat, the deletion may have resulted from recombination between the repeat sequences.

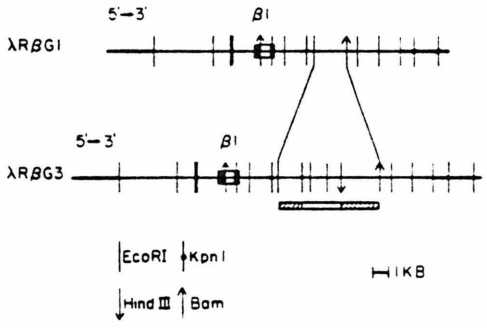

Figure 8. Mapping of a Deletion in λRβG1

The restriction endonuclease cleavage maps of λRβG1 and λRβG3 are compared to show the location of a deletion in λRβG1 that probably occurred during the propagation of λRβG3. Restriction sites missing in λRβG1 lie between the Eco RI and Bam HI sites connected by lines between the two maps. All other restriction sites in λRβG1 and λRβG3 are identical. The rectangular box beneath the λRβG3 map represents a 5.3 kb region from which 4.35 kb of DNA have been deleted to generate λRβG1 from λRβG3. The breakpoints of the deletion have not been mapped precisely, but the hatched areas locate the regions within which the breakpoints must occur. The clear area indicates the region of λRβG3 known definitely to be missing in λRβG1. Portions of the arms of the λCh4A vector are indicated by the thick horizontal lines. The restriction fragments missing in λRβG1 but present in λRβG3 are (Eco RI) 2.0, 1.1 and 4.0 kb; (Bam HI) 9.7 kb; (Kpn I) 1.95 and 9.25 kb. These are replaced by a 2.75 kb Eco RI fragment, a 5.35 kb Bam HI fragment and a 6.82 kb Kpn I fragment.

Discussion

We established the linkage arrangement and transcriptional orientation of four different β-like globin gene sequences and derived a detailed restriction map of the gene cluster. The transcriptional orientation and intergene distances are 5'-β4-8 kb-β3-5 kb-β2-7 kb-β1-3', placing β1, the adult β-globin gene, at the 3' end of the cluster. Information regarding the structure and transcription of all four genes is presented in the following paper (Hardison et al., 1979).

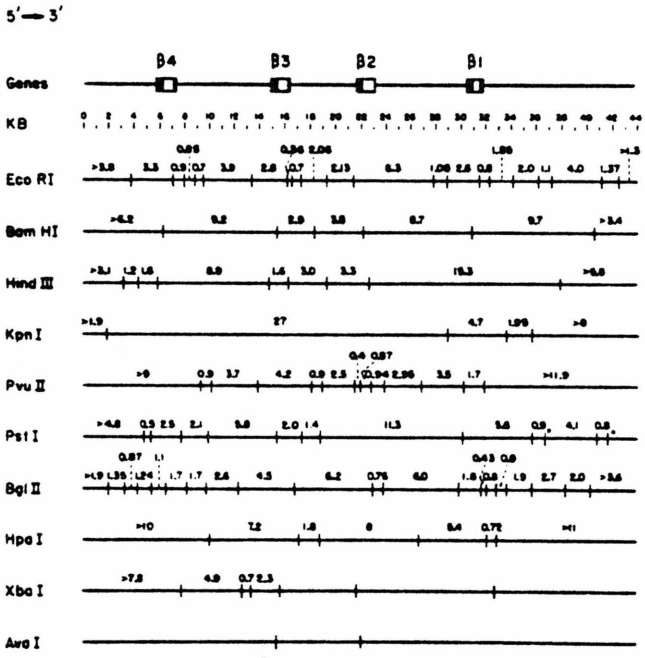

Figure 9. Restriction Endonuclease Cleavage Map of the Rabbit β-Like Globin Gene Cluster

Information from the restriction enzyme cleavage maps of the isolated genomic β-like globin gene clones was combined to generate the map shown here. The positions of the terminal restriction sites in rabbit DNA have not been determined. Asterisks indicate restriction sites which have not been located precisely. The Xba I and Ava I maps are not complete.

The isolation of 44 kb of contiguous rabbit chromosomal DNA was made possible by the development of a gene isolation procedure involving the construction of bacteriophage λ libraries of random genomic DNA fragments. These collections of recombinant phage are sufficiently large to have a 99% probability of containing any given region of genomic DNA (Maniatis et al., 1978). The utility of this approach to gene isolation is graphically illustrated in Figure 5, which shows that a large number of overlapping DNA fragments can be obtained by screening a library with a hybridization probe containing only 500 bp of sequence. Although one of the clones underwent a rearrangement, the correspondence between restriction sites in total and cloned genomic DNA in the rabbit β-globin gene cluster demonstrates the feasibility of isolating large, intact chromosomal DNA by molecular cloning procedures.

Other examples of close physical linkage between functionally related eucaryotic structural genes have been reported. These include the histone genes of sea urchin (for review see Kedes, 1979) and Drosophila (Lifton et al., 1978). In addition, three closely linked chicken ovalbumin-like genes have been isolated on two segments of DNA which together span 46 kb of the chicken chromosome (Royal et al., 1979). The organization of these genes is similar to that of the β-globin genes in that all of the ovalbumin-like genes are transcribed from the same DNA strand and share sequence homology. While all three ovalbumin-like genes appear to be coordinately expressed under hormonal control, the globin gene cluster consists of genes which are coordinately as well as sequentially expressed during development. Another example of linked genes which are transcribed from the same DNA strand are the genes encoding the 70 kb polypeptide induced by heat shock in Drosophila (Livak et al., 1978; Ashburner and Bonner, 1979; Craig, McCarthy and Wadsworth, 1979; Artavanis-Tsakonas et al., 1979; Moran et al., 1979). In addition to the genes which have been shown to be physically linked, a number of genes have been shown to be linked genetically. For example, genetic linkage has been demonstrated in complex loci such as the mouse immunoglobulin gene loci (Milstein and Munro, 1970; Hood, Campbell and Elgin, 1975) and the mouse (Klein, 1979) and human (Franke and Pellegrino, 1977) histocompatibility loci.

At the present time it is not known whether close physical linkage is merely the result of gene duplication and divergence with no regulatory significance or whether gene clusters have evolved as functional developmental units which include cis-acting regulatory sequences and genes which encode regulatory proteins or RNA. The clustering of certain genes must have functional significance; for example, the efficiency of somatic rearrangement of immunoglobulin genes during lymphocyte development is undoubtedly enhanced by physical linkage of variable and constant region genes (Tonegawa et al., 1978). It is also clear,

Cell
1280

however, that gene linkage is not a prerequisite for coordinate gene expression since mammalian α– and β–like globin genes, which are co-expressed in adult erythroid cells, are located on different chromosomes.

Regulatory interactions between linked sequences located many kilobases from each other are suggested by the fact that sequences found near the human δ– and β–globin genes can apparently affect the normal switch from fetal to adult globin gene expression (Huisman et al., 1974; Fritsch et al., 1979). The question of whether the linkage arrangement of mammalian β–like globin genes is functionally significant must await further studies.

Experimental Procedures

Isolation and Propagation of λCh4A Recombinant Clones

The construction and screening of a library of rabbit genomic DNA, the propagation and purification of λCh4A recombinant phage and the purification of phage DNA are described by Maniatis et al. (1978).

Digestion of DNA with Restriction Endonucleases and Electrophoresis

Restriction endonucleases were purchased from New England BioLabs or Boehringer-Mannheim. DNA was digested at 37°C in either of two buffers, 20 mM HEPES (Calbiochem) (pH 7.5), 30 mM NaCl, 6 mM $MgCl_2$, 6 mM dithiothreitol, 0.01% bovine serum albumin, supplemented with 70 mM NaCl (final 100 mM) for Ava I, Bam HI, Eco RI, Sal I and Xba I or with 20 mM KCl for Hpa I; or buffers recommended by New England BioLabs. The digested DNA was fractionated by agarose gel electrophoresis (Sharp, Sugden and Sambrook, 1973) using a horizontal slab gel apparatus.

Genomic DNA was digested with 1 unit of restriction endonuclease (1 unit of enzyme digests 1 μg of λ DNA in 1 hr in the New England BioLabs reaction conditions) per 3 μg DNA at 37°C for 12–16 hr. A side reaction containing 1 μg of genomic DNA and 0.4 μg of λ DNA was assayed for complete digestion by gel electrophoresis. The digests were terminated by extractions with phenol and ether, and the DNA was precipitated with ethanol, redissolved and electrophoresed.

Mapping Restriction Endonuclease Cleavage Sites

The restriction maps were determined by first digesting the cloned DNAs with either one or two restriction endonucleases and measuring the size of the resulting fragments on an agarose gel calibrated with restriction fragments from either pBR322 or ΦX174, the sizes of which are known from DNA sequence analysis (Sanger et al., 1977; Sutcliffe, 1978). Ambiguities in the double digest data were resolved by several different procedures. One method entailed purifying individual fragments from a single restriction enzyme digest by electroelution and then redigesting with a second enzyme. Electroelution involved either binding the DNA to hydroxyapatite on a preparative agarose gel (Tabak and Flavell, 1978) or simply electrophoresing the DNA into a trough in the gel filled with E buffer (Hayward and Smith, 1972), collecting fractions at appropriate times (determined by observing progress of electroelution by ultraviolet-induced fluorescence), passing the eluted DNA solution through siliconized glass wool, extracting with secondary butanol to remove ethidium bromide and to concentrate the DNA, and precipitating with ethanol. Alternatively, single restriction fragments were analyzed by a two-dimensional agarose gel electrophoresis procedure which involved the redigestion of fragments electrophoresed through low melting temperature "sea plaque" agarose gels (Lawn et al., 1978; Parker and Seed, 1979). In addition, subclones of restriction fragments from the recombinant phage clones were analyzed by combined restriction enzyme digests to obtain mapping information. Some cleavage sites were mapped using the Ca^{++}-dependent, double-stranded exonucleolytic activity in the extracellular nucleases from Alteromonas espejiana BAL31 (Legerski, Hodnett and Gray, 1978). The enzyme was provided by H. Gray.

Preparation of Plasmid DNA

Bacteria carrying plasmids were grown using high concentrations of uridine to enhance plasmid yield (Norgard, Emigholz and Monahan, 1979). Harvested cells were treated with 2 mg/ml lysozyme in 25% sucrose, 0.05 M Tris, 0.05 M EDTA for 10 min at 4°C, frozen in liquid N_2, thawed at room temperature and centrifuged at 60,000 × g, 30 min, 20°C to remove host chromosomal DNA (Conrad and Campbell, 1979). The supernatant, containing plasmid DNA, was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and twice with ether. Saturated CsCl and ethidium bromide were added to a final concentration of 0.3 mg/ml ethidium bromide, ρ = 1.60, and the solution was centrifuged at 40,000 rpm for 40 hr, 20°C in a 60Ti rotor. The supercoiled DNA band was collected, extracted twice with isopropanol equilibrated with saturated CsCl to remove ethidium bromide and dialyzed against 10 mM Tris–HCl (pH 8), 1 mM EDTA to remove CsCl.

Construction of Plasmid Subclones Containing Fragments of Cloned Rabbit DNA

Eco RI fragments isolated by electroelution were ligated with Eco RI-cleaved pMB9 (Rodriguez et al., 1976) that had been treated with calf intestinal alkaline phosphatase (Worthington Biochemicals) to remove 5′ phosphates and thereby reduce intramolecular ligation of the vector (Ullrich et al., 1977). The phosphatase was purified by passage over Sephadex G200 (Y. Chaconas, personal communication) prior to use. Bgl II fragments were ligated with dephosphorylated Bam HI–cleaved pBR322 DNA (Bolivar et al., 1977), and Pst I fragments were ligated with Pst I–cleaved pBR322 DNA that had not been treated with phosphatase. The Eco RI and Bgl II fragments were mixed with equimolar amounts of the appropriate vector at a combined DNA concentration of 100 ng/μl in 66 mM Tris–HCl (pH 7.6), 10 mM $MgCl_2$, 1.0 mM ATP, 15 mM dithiothreitol and 0.2 mg/ml gelatin. 0.5 units of T4 DNA ligase (Bethesda Research Labs, PL Biochemicals or New England BioLabs) were added and the reaction was incubated at 12°C for 15 min, then diluted to 10 ng/μl with ligase buffer and incubated for another 4–16 hr at 12°C. The Pst I fragments were ligated in a 10 fold molar excess over the Pst-cleaved vector under reaction conditions otherwise identical to those for Eco RI and Bgl II fragments.

Prior to the revision of the NIH Guidelines for recombinant DNA work on December 22, 1978, all in vitro recombinants were introduced, by transformation, into the EK2 certified host χ1776 (Curtiss et al., 1976) in a P3 facility. After the Guidelines were revised, the clones containing the isolated β–like genes were reclassified to P2 + EK1 containment. Recombinant plasmids were then used to transform the EK1 host HB101 (Boyer and Roulland-Dussoix, 1969) in a P2 laboratory. Transformation of χ1776 followed a protocol provided by D. Hanahan (personal communication). HB101 was transformed as described by Mandel and Higa (1970) or Morrison (1977). Transformants were screened by differential drug selection (Bolivar et al., 1977) and by a modification of the Grunstein and Hogness (1975) colony hybridization procedure. Nitrocellulose replicas from a master plate were lysed on Whatman 3MM paper saturated with 0.1 N NaOH, 1.5 M NaCl and neutralized on filter paper saturated with 0.2 M Tris–HCl (pH 7.5), 2 × SSCP [1 × SSCP: 120 mM NaCl, 15 mM sodium citrate, 13 mM KH_2PO_4, 1 mM EDTA (pH 7.2); Benton and Davis, 1977], 0.75 M NaCl, respectively. Recombinants formed from phosphatase-inactivated vectors gave up to 90% positives when transformants were screened. Only seven positives were found in 500 transformants screened (1.4%) for the Pst 2.5 kb subclone containing β4.

Blot Hybridization Experiments

DNA was transferred from agarose gels to Millipore HAWP 00010 or Schleicher and Schuell BA85 (0.45 μm) nitrocellulose filters and hybridized with radioactive probes (Southern, 1975) using the modifications of Jeffreys and Flavell (1977a) or Engel and Dodgson (1978). We currently follow the procedures of Wahl, Stern and Stark (1979), using dextran sulfate to accelerate the hybridization (Wetmur, 1975).

Table 1. Clones of Recombinant DNA Referred to in the Text

Full Name[a]	Abbreviated Name	Source of Inserted DNA	Gene Sequences	Vector Cleavage Site	Insertion Technique	Reference
(A) Genomic DNA						
λCh4A-Rβ Gn	λRβGn	Rabbit genomic DNA, Hae III/Alu I partial digest	β1 (β2, β3)[b]	Eco RI	Eco RI linkers	Maniatis et al. (1978)
λCh4A-Rβ′Gn	λRβ′Gn	Rabbit genomic DNA, Hae III/Alu I partial digest	β3 and β4	Eco RI	Eco RI linkers	This paper
pMB9-Eco RI y[c] (λRβ′′Gn)	pRy	Eco RI fragment from λRβG3, λRβG2, λRβ′G8		Eco RI	Ligation into phosphatase-treated vector	This paper
	pR 0.8		3′ of β1			
	pR 6.3		β2			
	pR 2.8		5′ of β3			
	pR 3.3		5′ of β4			
pBR322-Eco RI/Bam HI y (λRβ′′Gn)	pRBy	Eco RI/Bam HI fragment from λRβG2 or λRβ′G8		EcoRI/ Bam HI	Ligation	This paper
	pRB 1.95		5′ of β1			
	pRB 0.65		ivs of β1			
	pRB 0.6		5′ of β2			
	pRB 0.88 (β3)		ivs of β3			
	pRB 0.88 (β4)		ivs of β4			
pBR322-Pst I 5.6 (λRβG2)	pPst 5.6	5.6 kb Pst I fragment from λRβG2	β1	Pst I	Ligation	This paper
pBR322-Pst I 2.5 (λRβ′G8)	pPst 2.5	2.5 kb Pst I fragment from λRβ′G8	β4	Pst I	Ligation	This paper
pBR322-Bgl II 4.5 (λRβ′G8)	pBgl 4.5	4.5 kb Bgl II fragment from λRβ′G8	β3	Bam HI	Ligation into phosphatase-treated vector	This paper
(B) cDNA						
pMB9-Rcβ1[d]	pRcβ1	cDNA to rabbit adult β-globin mRNA	β1	Eco RI	AT tails	Maniatis et al. (1976)
pMB9-Hc^{G}γ-151[e]	pHcγ-151	cDNA to human fetal γ-globin mRNA	$^{G}\gamma$	Eco RI	AT tails	Wilson et al. (1978)
pBR322-Rcβ4-50	pRcβ4	cDNA to rabbit β4 mRNA	β4	Pst I	AT tails	E. Butler (unpublished)
pBR322-Rcβ3-32	pRcβ3	cDNA to rabbit β3 mRNA	5′ of β3	Pst I	AT tails	E. Butler (unpublished)
pBR322-Rcan-37	pRcan-37	cDNA to rabbit blood island globin RNA	Unidentified	Pst I	AT tails	E. Butler (unpublished)

[a] The name of each clone gives the vector followed by a descriptive designation of the recombinant DNA insert: (R) rabbit, (H) human, (/G) β-globin, (n) number of the isolate, (λ) bacteriophage λ vector, (p) plasmid vector.

[b] All λRβGn contain gene β1. See Figure 6 for those that contain β2 and β3.

[c] y is the size in kb of the fragment isolated from a clone of genomic DNA. Those subcloned Eco RI fragments containing gene sequences are listed in the table. Fragments currently subcloned are: Eco RI 1.3, 1.37, 4.0, 1.1 and 3.7 from λRβG3; Eco RI 2.0, 1.85, 0.8, 1.08, 6.3 and 1.65 from λRβG2; Eco RI 2.8, 3.3 and 3.4 from λRβ′G8.

[d] Names of cloned cDNA copies of mRNA include the designation "c" followed by the type of mRNA (van den Berg et al., 1978), for example, pRcβ1 is a plasmid containing a cDNA copy of rabbit globin gene β1. Isolate numbers are included only when necessary to avoid confusion. pMB9-Rcβ1 was previously called pβG1 (Maniatis et al., 1976).

[e] Designated pJW151 by Wilson et al. (1978).

Cell
1282

The hybridization solution contains 100 μg/ml poly(A) (PL Biochemicals) and 0.1% sodium dodecylsulfate (SDS) in addition to the other components listed by Wahl et al. (1979). After hybridization, the filter is washed twice in 50% formamide, 2 × SSC, 0.1% SDS, 0.1% sodium pyrophosphate followed by three washes in 50% formamide, 0.3 × SSC, 0.1% SDS, 0.1% sodium pyrophosphate, unless otherwise indicated. Washes are for 20–30 min at 42°C or at 68°C in aqueous solutions of identical salts. The washed filter is exposed to Kodak XR-5 X-ray film, with a Corning Lighting-Plus Intensifier Screen in some cases.

Hybridization Probes

DNA was labeled in vitro by nick translation (Maniatis, Jeffrey and Kleid, 1975). Radioactive cDNA was prepared from rabbit globin RNA as described by Engel and Dodgson (1978). α-^{32}P-deoxynucleoside triphosphates (400 Ci/mmole) were purchased from Amersham. E. coli DNA polymerase I was purchased from Boehringer-Mannheim or made in the Division of Biology at Caltech (J. Mueller). DNAase I was purchased from Sigma. DNA polymerase from avian myeloblastosis virus (AMV reverse transcriptase) was provided by Dr. J. W. Beard and the Office of Program Resources and logistics (Viral Cancer Program, NIH).

Acknowledgments

We would like to acknowledge Richard Parker, Brian Seed, Jerry Dodgson and Doug Engel for advice and discussions, and A. Efstratiadis for helpful comments on the manuscript. We thank C. O'Connell for constructing some of the subclones and A. Cortenbach for preparing media and material. We also thank B. Forget for providing the human γ-cDNA plasmid and David Goldberg for purifying the calf intestinal alkaline phosphatase.

E. L., R. C. H. and T. M. were supported by an NIH graduate training grant to the California Institute of Technology, a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research and a Rita Allen Foundation Career Development Award, respectively. This research was funded by grants from the National Science Foundation and Northwest Area Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 14, 1979; revised October 1, 1979

References

Artavanis-Tsakonas, S., Schedl, P., Mirault, M.-E., Moran, L. and Lis, J. (1979). Cell *17*, 9–18.

Ashburner, M. and Bonner, J. J. (1979). Cell *17*, 241–254.

Benton, W. D. and Davis, R. W. (1977). Science *196*, 180–182.

Blattner, F., Williams, B., Blechl, A., Denniston-Thompson, K., Faber, H., Furlong, L., Grunwald, D., Kiefer, D., Moore, D., Schuman, J., Sheiden, E. and Smithies, O. (1977). Science *196*, 161–169.

Bolivar, F., Rodriguez, R., Greene, P., Betlach, M., Heyneker, H. and Boyer, H. (1977). Gene *2*, 95–113.

Boyer, H. W. and Roulland-Dussoix, D. (1969). J. Mol. Biol. *41*, 459–472.

Bunn, H. F., Forget, B. G. and Ranney, H. M. (1977). Human Hemoglobins (Philadelphia: W. B. Saunders Co.).

Conrad, S. E. and Campbell, J. L. (1979). Nucl. Acids Res. *6*, 3289–3303.

Craig, E. A., McCarthy, B. J. and Wadsworth, S. C. (1979). Cell *16*, 575–588.

Curtiss, R., III, Pereira, A., Hsu, J., Hull, C., Clark, E., Maturin, J., Goldschmidt, R., Moody, R., Inoue, M. and Alexander, L. (1976). In Micro 76—The Annual Report (Birmingham: University of Alabama, Department of Microbiology), pp. 96–106.

Efstratiadis, A., Kafatos, F. C. and Maniatis, T. (1977). Cell *10*, 571–585.

Engel, J. D. and Dodgson, J. B. (1978). J. Biol. Chem. *253*, 8239–8246.

Flavell, R. A., Kooter, J. M., De Boer, E., Little, P. F. R. and Williamson, R. (1978). Cell *15*, 25–41.

Forget, B., Cavallesco, J., deRiel, J., Spritz, R., Choudary, P., Wilson, J., Wilson, L., Reddy, V. and Weissman, S. (1979). In Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, *XIV*, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press), in press.

Franke, U. and Pellegrino, M. (1977). Proc. Nat. Acad. Sci. USA *74*, 1147–1151.

Fritsch, E., Lawn, R. and Maniatis, T. (1979). Nature *279*, 598–603.

Gilman, J. and Smithies, O. (1968). Science *160*, 885–886.

Grunstein, M. and Hogness, D. (1975). Proc. Nat. Acad. Sci. USA *72*, 3961–3965.

Hardison, R. C., Butler, E. T., III, Lacy, E., Maniatis, T., Rosenthal, N. and Efstratiadis, A. (1979). Cell *18*, 1285–1297.

Hayward, G. and Smith, M. (1972). J. Mol. Biol. *63*, 383–395.

Hood, L., Campbell, J. and Elgin, S. (1975). Ann. Rev. Genet. *9*, 305–353.

Hunsman, T. H. J., Schroeder, W. A., Efremov, G. D., Duma, H., Mladenovski, B., Hyman, T. B., Rachmilewitz, E. A., Bouver, N., Miller, A., Brodie, A., Shelton, J. R., Shelton, J. B. and Apell, G. (1974). Ann. NY Acad. Sci. *232*, 107–124.

Jeffreys, A. J. and Flavell, R. A. (1977a). Cell *12*, 429–439.

Jeffreys, A. J. and Flavell, R. A. (1977b). Cell *12*, 1097–1108.

Kafatos, F., Efstratiadis, A., Forget, B. and Weissman, S. (1977). Proc. Nat. Acad. Sci. USA *74*, 5618–5622.

Kedes, L. (1979). Ann. Rev. Biochem. *48*, in press.

Kitchen, H. and Brett, I. (1974). Ann. NY Acad. Sci. *241*, 653–670.

Klein, J. (1979). Science *203*, 516–521.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. and Maniatis, T. (1978). Cell *15*, 1157–1174.

Legerski, R., Hodnett, J. and Gray, H. B., Jr. (1978). Nucl. Acids Res. *5*, 1445–1464.

Lifton, R., Goldberg, M., Karp, R. and Hogness, D. (1978). Cold Spring Harbor Symp. Quant. Biol. *42*, 1047–1051.

Little, P. F. R., Flavell, R., Kooter, J., Annison, G. and Williamson, R. (1979). Nature *278*, 227–231.

Livak, K., Freund, R., Schweber, M., Wensin, P. and Meselson, M. (1978). Proc. Nat. Acad. Sci. USA *75*, 5613–5617.

Mandel, M. and Higa, A. (1970). J. Mol. Biol. *53*, 159–162.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Maniatis, T., Sim, G. K., Efstratiadis, A. and Kafatos, F. C. (1976). Cell *8*, 163–182.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. and Efstratiadis, A. (1978). Cell *15*, 687–701.

Mears, J. G., Ramirez, F., Leibowitz, D. and Bank, A. (1978). Cell *15*, 15–23.

Milstein, C. and Munro, A. (1970). Ann. Rev. Microbiol. *24*, 335–358.

Moran, L., Mirault, M.-E., Tissières, A., Lis, J., Schedl, P., Artavanis-Tsakonas, S. and Gehring, W. J. (1979). Cell *17*, 1–8.

Morrison, D. A. (1977). J. Bacteriol. *132*, 349–351.

Norgard, M. V., Emigholz, K. and Monahan, J. J. (1979). J. Bacteriol. *138*, 270–272.

Parker, R. and Seed, B. (1979). In Methods in Enzymology, L. Grossman and K. Moldave, eds. (New York: Academic Press), in press.

Pirrotta, V., Ptashne, M., Chadwick, P. and Steinberg, R. (1971). In Procedures in Nucleic Acid Research, 2, G. L. Cantoni and D. Davies,

eds. (New York: Harper and Row), pp. 703–724.

Robinson, L. H. and Landy, A. (1977). Gene 2, 1–31.

Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. and Betlach, M. (1976). In Molecular Mechanisms in the Control of Gene Expression, ICN-UCLA Symposium on Molecular and Cellular Biology, 5, D. P. Nierlich, N. J. Rutter and C. F. Fox, eds. (New York: Academic Press), pp. 471–477.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J., Le Meur, M. Bregegegre, F., Gannon, F., LePennec, J., Chambon, P. and Kourilsky, P. (1979). Nature 279, 125–132.

Russell, E. S. and McFarland, E. C. (1974). Ann. NY Acad. Sci. 241, 25–38.

Sanger, F., Air, G., Barell, B., Brown, N., Coulson, A., Fiddes, J., Hutchison, C., III, Slocombe, P. and Smith, M. (1977). Nature 265, 687–695.

Sharp, P. A., Sugden, B. and Sambrook, T. (1973). Biochemistry 12, 3055–3063.

Southern, E. M. (1975). J. Mol. Biol. 98, 503–517.

Sutcliffe, J. G. (1978). Nucl. Acids Res. 5, 2721–2728.

Tabak, H. F. and Flavell, R. A. (1978). Nucl. Acids Res. 5, 2321–2332.

Tonegawa, S., Brack, C., Hozumi, N. and Pirrotta, V. (1978). Cold Spring Harbor Symp. Quant. Biol. 42, 921–931.

Tuan, D., Biro, A., deRiel, J., Lazarus, H. and Forget, B. (1979). Nucl Acids Res. 6, 2519–2544.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. and Goodman, H. (1977). Science 196, 1313–1319.

van den Berg, J., van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R. and Weissmann, C. (1978). Nature 276, 37–44.

Wahl, G. M., Stern, M. and Stark, G. R. (1979). Proc. Nat. Acad. Sci. USA 76, 3683–3687.

Weatherall, D. J. and Clegg, J. B. (1972). The Thalassemia Syndrome, 2nd edition (Oxford: Blackwell).

Wetmur, J. G. (1975). Biopolymers 14, 2517–2524.

Wilson, J. T., Wilson, L. B., deRiel, J., Villa-Komaroff, L., Efstratiadis, A., Forget, B. and Weissman, S. M. (1978). Nucl. Acids Res. 5, 563–581.

Chapter 3

Cell, Vol 18, 1285–1297, December 1979, Copyright © 1979 by MIT

The Structure and Transcription of Four Linked Rabbit β-Like Globin Genes

Ross C. Hardison, Eugene T. Butler, III, Elizabeth Lacy and Tom Maniatis
Division of Biology
California Institute of Technology
Pasadena, California 91125
Nadia Rosenthal and Argiris Efstratiadis
Department of Biological Chemistry
Harvard Medical School
Boston, Massachusetts 02115

Summary

Rabbit chromosomal DNA contains a cluster of four linked β-like globin genes arranged in the orientation 5′-β4-(8 kb)-β3-(5 kb)-β2-(7 kb)-β1-3′. Determination of the nucleotide sequence of gene β1 confirms that this gene corresponds to the second type of two common co-dominant alleles encoding the adult β-globin chain. With the exception of two nucleotide substitutions in the large intervening sequence (intron), the intron and flanking sequences are identical with the nucleotide sequence of the first type determined by Weissmann et al. (1979). A 14S polyadenylated transcript containing large intron sequences (possibly a mRNA precursor) is detected in the bone marrow cells of anemic rabbits. Gene β2 has limited sequence homology to adult and embryonic β-globin probes and lacks a detectable mRNA transcript in the erythropoietic tissues examined. It contains at least one intervening sequence analogous to the large intron in gene β1. Genes β3 and β4 both contain an intron of 0.8 kb. Partial DNA sequence analysis indicates that the large intron in β4 is located between codons for amino acids lysine and leucine in an analogous position to that of the large intron in β1. In addition, a second smaller intron interrupts the 5′ coding sequences of gene β4. Both genes β3 and β4 are transcribed in embryonic globin-producing cells. Their DNA sequence homology is limited, however, to a segment of approximately 0.2 kb located on the 5′ side of the large intron.

Introduction

We have described the isolation of a cluster of four different rabbit β-like globin gene sequences (β1–β4) which includes the adult β-globin gene β1 (Lacy et al., 1979). The cluster was obtained from a bacteriophage λ library of rabbit chromosomal DNA (Maniatis et al., 1978) as a set of overlapping phage recombinants which together include 44 kb (kilobase pairs) of contiguous DNA. As shown in Figure 1, the four genes are separated from each other by 5–8 kb of DNA and are transcribed from the same DNA strand in the orientation 5′-β4-β3-β2-β1-3′. It is not known whether this cluster comprises the entire rabbit β-globin gene family. In fact, relatively little is known about the number of different β-like globin polypeptides encoded in the rabbit genome or the timing of their expression during development. The information currently available is summarized in Table 1, which also includes a list of the human and mouse globin chains for comparison. Two types of embryonic globin chains, designated χ (or ξ) and ε, are synthesized in nucleated erythroid cells derived from the yolk sac blood islands of the embryo (Marks and Rifkind, 1972; Kitchen and Brent, 1974; Melderis, Steinheider and Ostertag, 1974; Steinheider, Melderis and Ostertag, 1975; Bunn, Forget and Ranney, 1977). The embryonic α-like globins, rabbit and mouse χ and human ξ, are produced early in development and are gradually replaced by the adult α-globin chains (Capp, Rigas and Jones, 1970; Melderis et al., 1974). Structural studies of rabbit and mouse ε-globin polypeptides suggest that they are encoded in at least two different ε-globin genes (Marks and Rifkind, 1972; Steinheider et al., 1975). They are replaced by adult β-globin polypeptides as the site of erythropoiesis shifts from the blood islands to the fetal liver (Kitchen and Brett, 1974). An ε-globin chain identified in human embryos (Huehns et al., 1964) shares sequence homology with both the human β- and γ-globin polypeptides (Gale, Clegg and Huehns, 1979), and is replaced by fetal γ-globin chains during the shift to fetal liver erythropoiesis. The γ-globin polypeptides are replaced in turn by the adult β- and δ-globins at the time of birth (Weatherall and Clegg, 1979). Thus the current view is that two switches in β-like globin gene expression (ε → γ → β) take place during human development, while only one is observed in rabbit and mouse.

The available polypeptide data indicate that the rabbit and mouse globin gene families consist of at least three or four β-like genes and two α-like genes. More genes may yet be identified, however, since these data provide only a minimal estimate. In this paper we report structural and transcriptional analyses of each of the four isolated rabbit β-like globin genes. We find that these genes have structural features in common with each other and with globin genes in other species. Moreover, our studies show that the isolated genes are differentially expressed during development and suggest the presence of additional β-like genes in the rabbit genome.

Results

Structure of Gene β1

Gene β1 was initially identified as the adult β-globin gene on the basis of blot hybridization experiments using the adult β-globin cDNA plasmid pRcβ1 (Maniatis et al., 1976; for nomenclature see Lacy et al.,

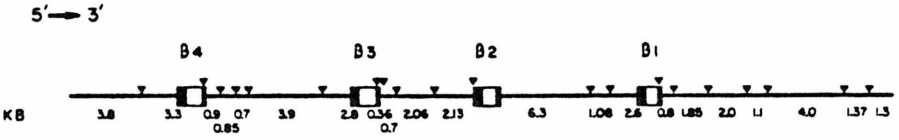

Figure 1. Linkage Map of Four Rabbit β-Like Globin Genes

The locations of Eco RI cleavage sites are indicated by inverted triangles (see Lacy et al., 1979 for a more detailed restriction map). The direction of transcription of all four genes is from left to right. The black and white boxes indicate the mRNA coding and noncoding sequences, respectively.

Table 1. Mammalian Globin Polypeptide Chains

		Embryonic	Fetal	Adult
Rabbit	α-like	χ		α
	β-like	ε(γ), ε(2)		β
Mouse	α-like	χ		α
	β-like	ε(γ), ε(2)		β-minor, β-major
Human	α-like	ζ		α_1, α_2
	β-like	ε	$^{G}\gamma$, $^{A}\gamma$	δ, β

1979) as the hybridization probe (Maniatis et al., 1978; Lacy et al., 1979). These experiments demonstrated that the restriction sites in the flanking regions of the cloned β1 gene were identical to those surrounding the adult β-globin gene in rabbit chromosomal DNA (Jeffreys and Flavell, 1977a). A comparison of the restriction map of pRcβ1 (Efstratiadis, Kafatos and Maniatis, 1977) with the map of sites within the corresponding sequences in total rabbit genomic DNA revealed the presence of a large (600 bp) noncoding intervening sequence within the rabbit β-globin gene (Jeffreys and Flavell, 1977b). A similar interruption was independently and simultaneously discovered within the adult mouse β-major globin gene by analysis of the structure of R loops formed between mouse β-globin mRNA and cloned chromosomal DNA (Tilghman et al., 1978a). The large intervening sequence was localized between the codons for amino acids 101 and 120 in the rabbit (Jeffreys and Flavell, 1977b) and 104 and 105 in the mouse (Tilghman et al., 1978a) β-globin genes. A second, smaller intervening sequence was subsequently identified between the codons for amino acids 30 and 31 in both the mouse (Konkel, Tilghman and Leder, 1978) and rabbit (van den Berg et al., 1978) β-globin genes. Thus the mature β-globin mRNAs are encoded in three discontinuous blocks separated by two noncoding intervening sequences (IVS or introns). To characterize gene β1 fully and to study its transcription, we constructed a detailed restriction map of clone λRβG1 and sequenced the regions containing the gene by using the procedure of Maxam and Gilbert (1977). In the mature rabbit β-globin mRNA there are 589 nucleotides (nt) between the sites to which the cap and poly(A) are attached (Efstratiadis et al., 1977). In the gene there are 1288 bp of DNA between the corresponding sites (Figure 2). The following five segments are present within this region: exon 1 (146 bp), which includes a region corresponding to the 5′ untranslated sequence of the mature mRNA (53 bp, excluding the initiation codon) and 93 bp of translated sequence; intron I (small intron, 126 bp); exon 2 (222 bp of coding sequence); intron II (large intron, 573 bp); and exon 3 (221 bp), which includes the remaining coding sequence (129 bp, including the termination codon) and the region corresponding to the 3′ untranslated sequence (93 bp) of the mature mRNA (Figure 3). The sequence of β1 corresponds to the second type of two common co-dominant rabbit β-globin alleles (Galizzi, 1970, 1971; Shamsuddin et al., 1973; Bricker and Garrick, 1974; Garrick et al., 1974) which show sequence polymorphism at amino acid positions 52, 56, 76 and 112. These two types are designated 1 (52 Asn, 56 Asn, 76 Ser, 112 Ile) and 2 (52 His, 56 Ser, 76 Asp, 112 Val). The nucleotide sequence of the β-globin cDNA plasmid pRcβ1 (Efstratiadis et al., 1977) and the sequence of chromosomal rabbit β-globin DNA determined independently from the plasmid Rchrβ G1 (van den Berg et al., 1978; Mantei et al., 1979; Weissmann et al., 1979) correspond to the type 1 allele. Comparison of the λRβG1 and Rchrβ G1 sequences reveals that the four amino acid replacements result from single base substitutions. It is interesting that silent substitutions are completely absent, which suggests the existence of selective constraints for the preservation of the secondary structure of the mRNA and/or its precursors (see also Kafatos et al., 1977). The introns of the rabbit β-globin gene are at the same positions as in the corresponding gene of the mouse (Konkel et al., 1978). The small intron has a four nucleotide terminal redundancy at its boundaries with the exons while the large intron has only one. Thus, as in the case of the mouse β-globin introns (Konkel et al., 1978; van den Berg et al., 1978) and other introns (for reviews see Crick, 1979; Dawid and Wahli, 1979), a unique splicing frame cannot be defined by the sequence data. The generalization that the base sequence of an intron begins with GT and ends with AG (Breathnach et al., 1978) is obeyed in only one splicing frame for each intervening sequence (see Figure 2). The splicing junctions show little deviation from the prototype sequences TCAGGTA (5′ end) and TXCAGG (3′ end) identified by Breathnach et al. (1978) and Catterall et al. (1978).

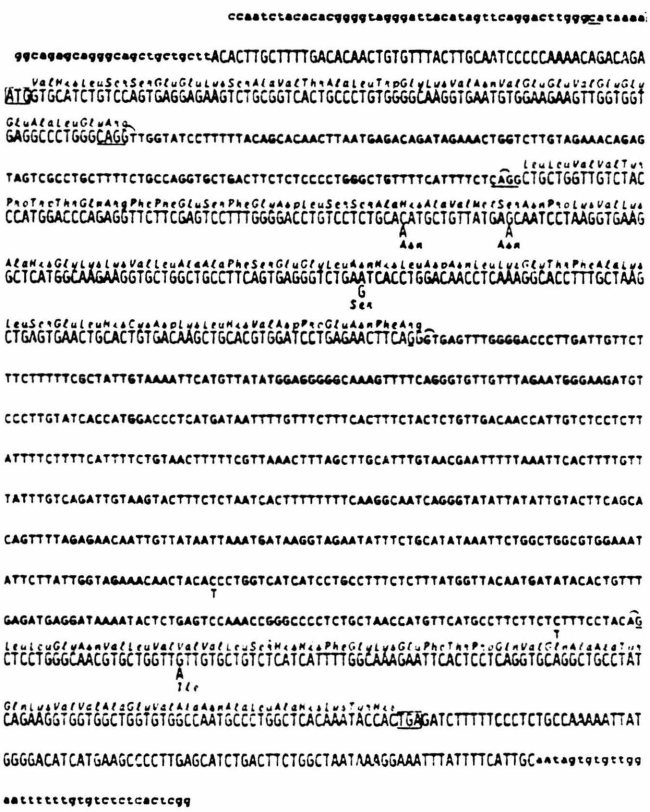

Figure 2. Nucleotide Sequence of the Rabbit Adult β-Globin Gene (β1)

The nucleotide sequence of the message strand is displayed from the 5′ to the 3′ direction. Nucleotide substitutions in the other known allele of this gene (Efstratiadis et al., 1977; van den Berg et al., 1978; Mantei et al., 1979; Weissmann et al., 1979) are indicated in a second line. Large and small capital letters represent sequences corresponding to mature mRNA and the introns, respectively. Gene flanking sequences are in lower case letters. The amino acid sequence is displayed on a line above the coding sequence. The initiation and termination codons and a sequence found 23 (±1) nucleotides 5′ to the cap site in the flanking sequences of many eucaryotic genes are boxed. Nucleotide redundancies at the intron–exon junctions are underlined, and the GT and AG dinucleotides at the beginning and end of the introns are overlined.

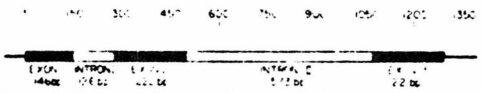

Figure 3. Schematic Diagram Showing the Distribution of mRNA Coding (Exon), Noncoding (Intron) and Flanking Sequences in Gene β1

With the exception of two nucleotide substitutions in the large intron (both C → T transitions), both intervening sequences are in complete agreement with the DNA sequence reported for Rchrβ G1 (van den Berg et al., 1978; Mantei et al., 1979; Weissmann et al., 1979). The same is true for the gene flanking sequences determined thus far. As in the case of other eucaryotic genes (Ziff and Evans, 1978; Gannon et al., 1979; Lomedico et al., 1979), a variant of the sequence TATAAATA first described by Goldberg and Hogness (Goldberg, 1979), CATAAAA(G) in gene β1, is present in the 5′ gene flanking region, 24 nucleotides upstream from the capping site. Although the position of this sequence is suggestive, there is no evidence that it is related to the initiation of transcription. Another sequence GCTGCTGCTTAC which overlaps the cap site is similar to sequences found in the same region of a number of mammalian cellular (Konkel et al., 1978) and viral (Ziff and Evans, 1978) genes. A detailed evolutionary comparison of the rabbit β-globin gene and other mammalian globin genes will be presented elsewhere.

Transcription of Gene β1

15S mRNA precursors of the adult mouse β^{maj} and β^{min} globin genes have been reported (Ross, 1976; Curtis and Weissmann, 1976; Bastos and Aviv, 1977; Kwan, Wood and Lingrel, 1977; Haynes et al., 1978). In addition to the sequences present in the mature mRNAs, the precursors contain both intervening sequences arranged co-linearly with the chromosomal gene (Smith and Lingrel, 1978; Kinniburgh, Mertz and Ross, 1978; Tilghman et al., 1978b). To determine whether an analogous precursor of gene β1 is present in rabbit erythroblasts, a blot of poly(A)-containing RNA from the bone marrow of an anemic rabbit (Alwine, Kemp and Stark, 1977) was hybridized to a ^{32}P-labeled plasmid (pRBO.65) which contains the large IVS from gene β1 plus 67 bp of coding sequence.

Cell
1288

Genomic blotting experiments indicate that the $\beta 1$ large intron is single copy, so any RNA species identified using it as a probe must be transcripts of gene $\beta 1$. Three RNAs of 1400, 900 and 675 nt (±50 nt) were detected with this probe (Figure 4). The 1400 nt species presumably corresponds to a polyadenylated transcript containing 1288 nt encoded in gene $\beta 1$. The 675 nt RNA is the size expected for the polyadenylated mature globin mRNA. The hybridization of this species is probably due to the presence of 67 bp of coding sequence in the probe. Although the blot was washed at a stringency expected to melt a 67 bp RNA-DNA duplex (see legend to Figure 4), there is probably as much as 10^3 times more mRNA than precursor RNA in immature adult erythroid cells [This is based on estimates of 10,000–24,000 mRNA molecules per cell in rabbit bone marrow (Clissold, Arnstein and Chesterton, 1977) and 17–85 precursor molecules per mouse erythroid cell (Ross and Knecht, 1978).] Thus as little as 0.1% residual counts remaining after washing the filter would produce an mRNA signal equal to that of the precursor. The 900 nt RNA species may represent a processing intermediate containing sequences from the large intron, but further studies are required to confirm this possibility. Such an intermediate would be consistent with a processing pathway involving a two step removal of the large intron proposed for the mouse β-globin precursor RNA (Kinniburgh and Ross, 1979). Transcriptional studies of the rabbit β-globin gene using an S1 nuclease mapping procedure (Berk and Sharp, 1977) indicate that the 5′ and 3′ termini of the large globin RNA precursor are encoded at the same positions as the termini of mature globin mRNA (Flavell et al., 1979), but there is no direct evidence that this precursor is the primary transcript. This analysis of processing intermediates also reveals a stepwise removal of the intervening sequences, the first step being excision of the small intron. The processing pathway for the large intervening sequence has not yet been definitively established.

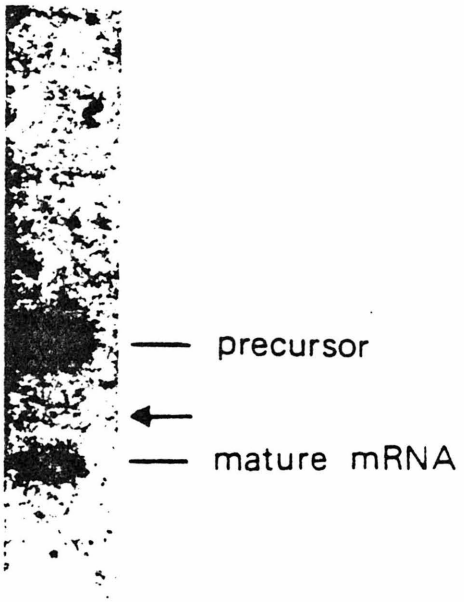

Figure 4. Detection of $\beta 1$ Transcripts in Bone Marrow RNA

An autoradiogram is shown of an RNA blot hybridized with a ^{32}P-labeled plasmid subclone containing the large intron of gene $\beta 1$ plus 18 and 49 bp of exon 1 and exon 2, respectively. Bands corresponding to RNAs of approximately 1400, 900 and 675 nt are indicated. 20 μg of total poly(A) RNA from the bone marrow of anemic adult rabbits were reacted with glyoxal and run in a single slot of a 1.5% agarose gel (McMaster and Carmichael, 1977). The RNA was transferred to diazotized paper (B. Seed, personal communication) as described by Wahl et al. (1979) and hybridized to 1 μg of nick-translated pRB0.65 DNA (spec. act 8.5 × 10^7 cpm/μg) in 20 ml. After hybridizing for 36 hr, the blot was washed 4–5 times in 50% formamide, 1 × SSC and 0.1% SDS at 45°C. A final wash was performed in 0.2 × SSC and 50% formamide for 30 min. Under these conditions, an RNA-DNA hybrid of 70 bp should have a T_m of 37°C while an RNA-DNA hybrid of 600 bp should be stable (J. Casey, personal communication). The blot was exposed to X-ray film for 5 days at −70°C using a DuPont Lightning Plus intensifying screen. The 28S and 18S ribosomal RNA bands and the 9S mature globin mRNA band, visible on the ethidium bromide-stained gel, were used as size standards.

Hybridization of $\beta 2$, $\beta 3$ and $\beta 4$ to Adult and Embryonic RNAs

To obtain preliminary information on the identities of genes $\beta 2$, $\beta 3$ and $\beta 4$, we examined their hybridization to probes for adult, fetal and embryonic globin RNA sequences. When the rabbit β-globin cDNA plasmid pRc$\beta 1$ (adult β-globin sequence) is used to probe a blot of Eco RI fragments containing all four cloned genes (Figure 5A), the fragments containing the 5′ and 3′ $\beta 1$ sequences display intense hybridization signals, as expected. $\beta 2$, 5′$\beta 3$ and 5′$\beta 4$ fragments hybridize weakly; the 3′$\beta 3$ fragment is very faint and the 3′$\beta 4$ fragment is not detected. Since gene $\beta 2$ shows weak homology with gene $\beta 1$, it is unlikely to be analogous to the human minor adult δ-globin gene, which shares 93% amino acid sequence homology (Dayhoff, 1972) and hybridizes efficiently to the human adult β-globin probe (Lawn et al., 1978). In contrast, when the same set of gene fragments is hybridized to the human γ-globin cDNA plasmid pHcγ-151 (fetal globin sequences), the 5′$\beta 3$ and 5′$\beta 4$ fragments hybridize as efficiently as the 5′$\beta 1$ fragment and display stronger signals than the $\beta 2$ fragment. 3′

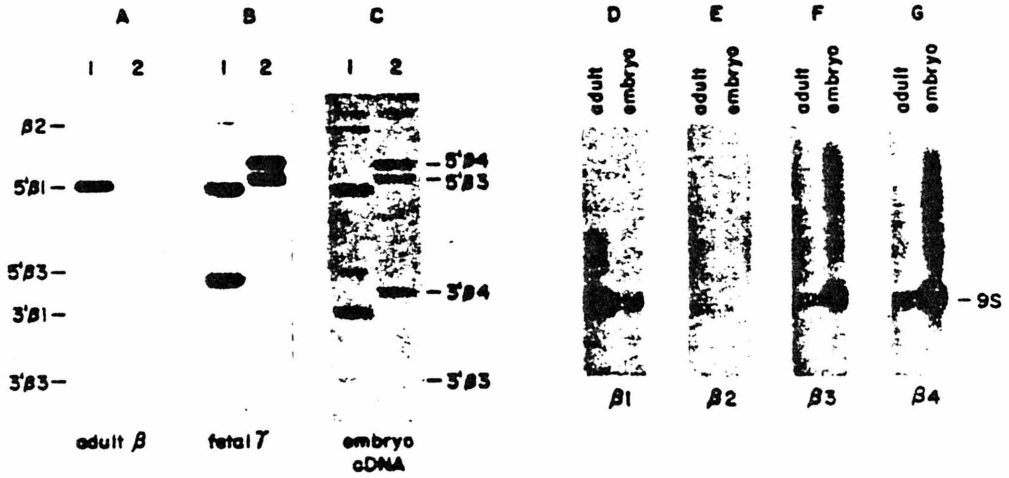

Figure 5. Preliminary Identification of Genes β1 → β4: Hybridization Analysis

(A–C) Cloned DNA fragments containing all four gene sequences were hybridized to various β-like globin cDNA probes using the blot hybridization procedure. DNAs from clone λRβG5 (lane 1) or λRβ'G8 (lane 2) (0.2 μg each) were digested with Eco RI, fractionated on a 1.5% agarose gel, blotted and hybridized with (A) pRcβ1 (adult rabbit β-globin probe); (B) pHcγ-151 (human fetal γ-globin probe); (C) cDNA synthesized from cytoplasmic poly(A) RNA obtained from reticulocytes from 12 day rabbit embryos. Autoradiograms of the blots are shown. The band visible above the β2 fragment in (C) is a λCh4A arm; this hybridization is not reproducible.

(D–G) Blots of adult erythroid and fetal embryonic RNA were hybridized to probes specific for each of the genes β1–β4. 0.8 μg of total poly(A) RNA from the bone marrow of anemic adult rabbits (adult) or 4 μg of total poly(A) RNA from 12 day rabbit embryos (embryo) were fractionated on a 1% agarose, 6 M urea gel (Lehrach et al., 1977). The two lanes contain equivalent amounts of 9S RNA. The RNA was transferred to diazotized paper and hybridized as described in Experimental Procedures. Autoradiograms are shown. The probes used were (D) β1: pRcβ1; (E) β2: an Eco RI/Bgl II 1.5 kb fragment from λRβG2; (F) β3: a Pvu II 4.2 kb fragment from λRβ'G8; (G) β4: a Pst I 2.5 kb fragment from λRβ'G8. The hybridization detected in the regions corresponding to RNAs larger than 9S could be due to the presence of precursor RNAs or nonglobin transcripts from gene-flanking regions, or to aggregation.

fragments from genes β3 and β4, however, are not detected by the γ probe (Figure 5B).

The stronger hybridization of genes β3 and β4 to the fetal globin sequence than to the adult sequence suggests that they are expressed during rabbit embryonic and/or fetal development. To test this possibility, we compared the hybridization of each of the four genes to embryonic cDNA. The cDNA was synthesized from cytoplasmic poly(A) RNA, containing mostly 9S material, from the blood islands of 12 day rabbit embryos. This RNA should contain transcripts derived primarily from the nucleated reticulocytes known to synthesize embryonic globins (Kitchen and Brett, 1974). The probe produced intense hybridization signals with the 5' and 3'β4 and 5'β3 fragments, as well as a faint but reproducible signal with the 3'β3 fragment. This result indicates that these genes have extensive homology with embryonic globin transcripts (Figure 5C). The embryonic cDNA also hybridizes efficiently to 5' and 3'β1 fragments but weakly to the gene β2 fragment, indicating the presence of adult β-globin transcripts in embryonic reticulocytes. The prevalence cannot be estimated, however, because the intensity of hybridization to β1 was variable with probes derived from a series of embryonic RNA preparations. In contrast, genes β3 and β4 displayed strong hybridization with all preparations of the embryonic probes.

This evidence for differential transcription of the β-like genes was complemented by hybridizing probes containing each of the four β-like sequences to RNA blots containing total poly(A) RNA from the bone marrow of anemic adult rabbits and from whole embryo (including blood islands). Figure 5D shows that gene β1 hybridizes efficiently to adult globin RNA, especially in the 9S region, but only weakly to embryonic RNA. Gene β2 hybridizes much less efficiently than gene β1 and exclusively with the adult RNA (Figure 5E). It is probable that this hybridization results from cross-hybridization to β1 transcripts, because we were unable to detect a transcript of gene β2 in the RNA preparations described in Figure 5. R loops were not observed between β2 and adult globin RNA under conditions identical to those that allow R loop formation in gene β1, even though the T_m for gene β2 DNA is comparable to the T_m for β1 (data not shown). In addition, we were unable to find a β2 transcript in adult or embryonic RNA using an S1 nuclease-mapping procedure (Berk and Sharp, 1977) which, under our conditions, can detect about ten copies of globin mRNA per cell (Wold et al., 1979). For technical reasons, however, we are unable to rule out the pos-

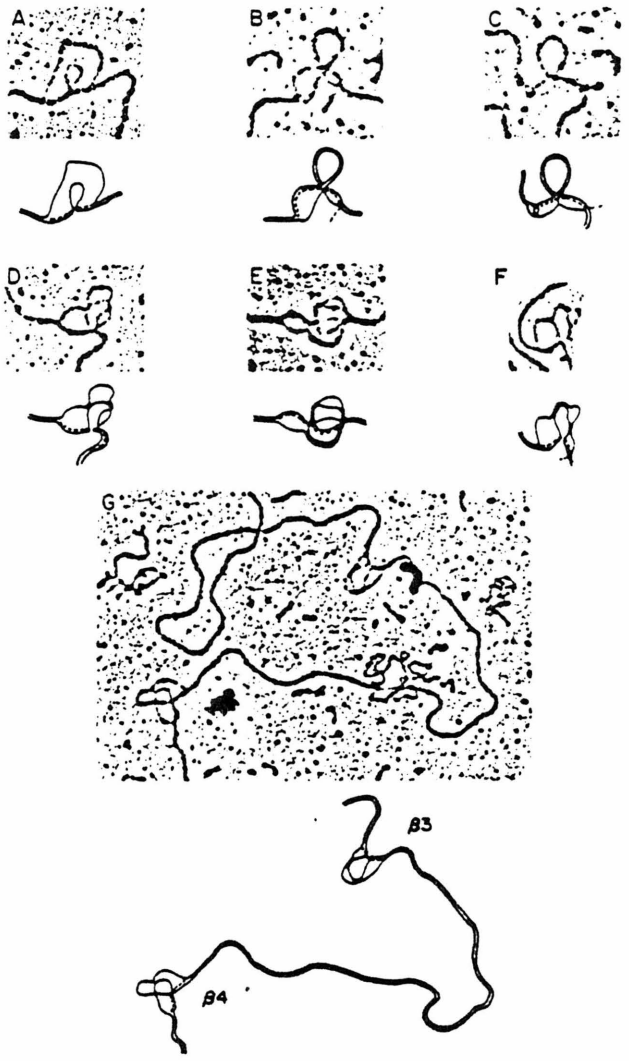

Figure 6. R Loop Analysis of Genes β3 and β4

Electron micrographs are displayed of R loops formed between total poly(A) RNA from whole 12 day rabbit embryos and λRβ'G8, containing β3 and β4 (A, D–G), or a 9.2 kb Pvu II fragment from λRβ'G8 containing gene β4 (B–C). Interpretative drawings are included beneath each photograph, with a solid line corresponding to a single strand of DNA and a dotted line denoting a strand of RNA. In (D), partial annealing has occurred between the two strands of the intervening sequence (denoted by cross-hatching), producing an apparent underwound loop (Broker et al., 1977). One of the coding halves of the gene has collapsed. The molecules shown in (E) and (F) have formed longer duplexes in the intervening sequence, leaving a single-stranded bubble between them. The tangle of single-stranded nucleic acid visible between β3 and β4 in (G) could be a coincidental overlap of RNA or possibly hybridization of a nonglobin sequence. The latter possibility is suggested by the fact that cDNA synthesized from total poly(A) embryonic RNA hybridizes to this region in blot hybridization experiments (results not shown).

sibility that β2 is transcribed into an unprocessed precursor. It is also possible that gene β2 is expressed in a tissue or developmental stage we have not examined.

In contrast to the β1 and β2 results, the β3 and β4 probes hybridize more efficiently to embryonic RNA than to adult RNA (Figures 5F and 5G), again predominantly in the 9S region. These data were quantitated by scanning the autoradiograms with a microdensitometer, and were normalized for the amount of 9S RNA present in each lane and the length of exposure. Gene β1 shows a 7 fold greater hybridization to adult globin RNA, whereas genes β3 and β4 show 8–17 fold greater hybridization to embryonic globin RNA. It is therefore probable that genes β3 and β4 encode embryonic β-like globin polypeptides, but these hybridization data do not constitute formal proof. As a means of obtaining more conclusive evidence for the transcription of these genes in embryonic erythroid cells, we have cloned cDNA synthesized from embryonic RNA. The cDNA clones were obtained using an mRNA/cDNA hybrid cloning procedure (Wood and Lee, 1976; Zain et al., 1979). The isolation and characterization of these clones will be described else-

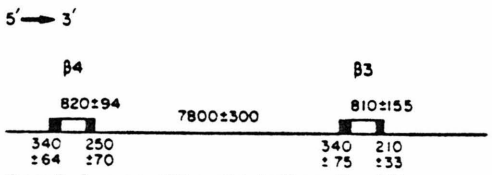

Figure 7. Summary of R Loop Data for Genes β3 and β4

This schematic drawing summarizes the structural features of genes β3 and β4 derived from measurements of 34 molecules. The average sizes (± standard deviation) are indicated in base pairs. The black boxes correspond to mRNA coding regions and the white boxes denote large, noncoding, intervening sequences within the genes.

where. Recombinant plasmids were selected according to their preferential hybridization to cloned genomic DNA containing β3 and β4. One of the plasmids (pRcβ3) hybridizes specifically to cloned genomic DNA containing β3 and another (pRcβ4) hybridizes specifically to β4. These two recombinant plasmids are not homologous to other chromosomal fragments, as demonstrated by genomic blotting analysis (not shown); thus they are not derived from other genes related to β3 or β4. This result indicates that genes β3 and β4 are transcribed in embryonic erythroid cells.

R Loop Analysis of β3 and β4

A large intervening sequence was observed in both genes β3 and β4 by analyzing R loops formed between embryonic RNA and cloned genomic fragments containing these genes (Kaback, Angerer and Davidson, 1979). The electron micrograph of Figure 6A shows that the formation of an RNA-DNA hybrid has displaced the message-synonymous strand in the gene region. A single-stranded segment of DNA on the message-complementary strand loops out in the middle of the region hybridized to the RNA. This looped-out segment of the complementary strand contains DNA not included in the mature mRNA and is therefore an intervening sequence separating the coding regions of the gene. Alternative forms of the R loop (Figures 6B and 6C) show the two strands of the intervening sequence as a complete DNA duplex separating the two halves of the R loop. A small knob visible in one side of the R loop in Figure 6C may correspond to an additional small intervening sequence analogous to that observed in the adult β-globin gene.

One technical aspect of the R loop analysis is worth noting. We frequently observed twisted structures such as those shown in Figures 6D–6G. Measurements indicate that these structures map to the β3 and β4 gene regions. This type of R loop probably forms when the single-stranded intervening sequence shown in Figure 6A hybridizes with the complementary sequence on the displaced strand. A complete duplex cannot form, however, because of topological constraints introduced by the RNA-DNA hybrid at the base of the intervening sequence loop. As the two strands wind in one direction, another portion of the intervening sequence must wind in the opposite direction to relieve the superhelical tension. These structures (Figures 6D–6F) are probably equivalent to underwound loops which are formed when DNA containing separated inverted duplications is denatured and reannealed (Davis, 1969; Berg et al., 1975; Kleckner et al., 1975; Broker, Soll and Chow, 1977).

Large intervening sequences in both genes β3 and β4 are shown in Figure 6G. The intervening sequences are centrally located in the coding region, and in this example have formed short hybrids with the complementary DNA strand. Measurements of 34 molecules established the sizes of the coding and noncoding regions of the two genes. In addition, the asymmetry in the lengths of the λCh4A arms (19.2 versus 10.7 kb) made it possible to align the R loop map with the restriction map (Lacy et al., 1979) and thereby determine the 5′ → 3′ orientation of the R loops. Whiskers are occasionally observed at the 3′ ends (Figures 6B, 6C, 6F) which presumably contain poly(A). A summary of the average values for the sizes of genes β3 and β4 is shown in Figure 7. Both genes have virtually identical dimensions, with a 5′ coding region of 340 ± 75 bp, a large intervening sequence of 815 ± 155 bp and a 3′ coding region of 230 ± 70 bp. The genes are separated by about 8 kb of DNA.

Structure of Gene β4

To determine the precise location of the large intervening sequence in gene β4, and to determine whether the gene encodes a known rabbit β-like globin polypeptide, we derived the nucleotide sequence flanking the intragenic Eco RI site in the β4 cDNA and genomic clones. These sequences are identical from codons 105 to 138 (Figure 8). The sequence of the genomic clone differs from the cDNA clone before codon 105, so the large intervening sequence terminates at codon 105 as in gene β1. As with the adult β-globin genes (Konkel et al., 1978; van den Berg et al., 1978), the sequence around the intron-exon junction is conserved between β1 and β4. Sixteen of the 26 nucleotides preceding the junction are identical in β1 and β4. Out of the 141 nucleotides determined for the protein-coding region of gene β4, the sequence differs from rabbit adult β-globin in 23 positions and from human γ-globin (Forget et al., 1979) in 18 positions (16 and 13% difference, respectively). The predicted amino acid sequence differs significantly from the rabbit ε-globin sequence reported by Steinheider et al. (1975). Twelve of the predicted 48 amino acid residues are different in this region (25% difference). The lack of correspondence between these sequences indicates that either β4 does not encode an ε-globin chain or that the published polypeptide sequence for the region covered by our DNA sequence is not correct.

```
                91          95          100         104
Rabbit ε
       β4   LeuHisCysAspLysLeuHisValAspProGluAsnPheLys
       β4   CTGCACTGTGACAAACTGCATGTGGATCCTGAGAACTTCAAA ... ~800 nt... ATATGCCTAT
                                      ↑
                                   Bam HI

                                   105         110         115
Rabbit ε                                  Ala   AlaThrCysAsxGlx   SerHis
       β4                            LeuLeuGlyAsnValLeuLeuIleValLeuAlaThrHisPhe
       β4   GTATCATTTTGTCTTTTACCTAACAGCTCCTAGGAAACGTGCTGCTGATCGTTCTGGCCACTCATTTT
       β1   TCATGCCTTCTTCTCTTTCCT-ACAGCTCCTG

                120         125         130         135     138
                (Glx,
                 Asx,                   (Ala,   (Leu, (Val,
Rabbit ε         Ser)       Gln         Ser)    Ile)  Ala)
       β4   GluLysGluPheThrProGluValGlnAlaAlaTrpGlnLysLeuValSerGlyValAla ...
       β4   GGCAAAGAATTCACTCCGGAGGTGCAGGCTGCTTGGCAGAAGCTCGTGTCTGGTGTTGCA ...
                ↑
             Eco RI
```

Figure 8. Partial Nucleotide Sequence of Gene $\beta 4$

The nucleotide sequence flanking the intragenic Eco RI site in $\beta 4$ cDNA and genomic DNA clones is shown along with the predicted amino acid sequence. The differences between the ϵ-globin sequence of Stemhelder et al. (1975) and the $\beta 4$ sequence are indicated. Due to heterogeneity in the ϵ-globin sequence, more than one amino acid is indicated at some positions. Large and small capital letters represent sequences corresponding to the mature mRNA and the large IVS, respectively. The numbers correspond to the amino acid positions in $\beta 1$.

The nucleotide sequence of the genomic $\beta 4$ DNA differs from the cloned cDNA at positions before the leucine codon at 105. This region, which contains the 3′ end of the large intervening sequence, is compared with the corresponding region of the $\beta 1$ gene. The nucleotide sequences were derived from the $\beta 4$ cDNA plasmid pRc$\beta 4$ and from a plasmid subclone containing a 2.5 kb Pst I fragment from λRβ′G8 containing the $\beta 4$ gene (pPst2.5). The sequence was not confirmed by sequencing the complementary strand.

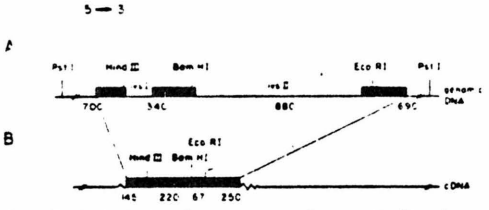

Figure 9. Evidence for a Small Intervening Sequence in Gene $\beta 4$

(A) A restriction map of the 2500 bp Pst I fragment containing gene $\beta 4$. This fragment, derived from λRβ′G8, was inserted into the Pst I site of pBR322 (Bolivar et al., 1977). Sequences found in mRNA are shown as black boxes separated by intervening sequences (ivs I and ivs II). The junction between the 3′ end of the large intervening sequence (ivs II) and the remainder of the mRNA-coding segment were determined by nucleotide sequence analysis (Figure 8). The other boundaries of the gene assume a direct analogy with the structure of gene $\beta 1$ (Figure 3). Only portions of the 5′ and 3′ flanking segments are shown. Sizes of DNA fragments are in base pairs.
(B) A restriction map of the cDNA insert in pRc$\beta 4$. The serrated line denotes the poly d(A)-d(T) tails used to insert the cDNA into the Pst I site of pBR322.

The identity of coding sequences derived from the $\beta 4$ genomic and cDNA clones argues strongly that the mRNA sequence represented in the cloned cDNA is in fact transcribed from the $\beta 4$ gene. This conclusion is supported by the observation that the $\beta 4$ cDNA clone detects only one fragment in a genomic blot and the size of this fragment is identical to that found in λRβ′G8. Alignment of restriction maps of the $\beta 4$ genomic and cDNA clones confirms the location and size of the large intervening sequence and reveals the presence of a second interruption in the $\beta 4$ gene. As shown in Figures 9A and 9B, the intragenic Bam HI and Eco RI sites are separated by 67 bp in the cDNA clone and by 880 bp in the genomic clone. The size of the large intervening sequence must therefore be 810 bp, a size in excellent agreement with that estimated by R loop analysis (Figure 7). The presence of a second intervening sequence of approximately 120 bp is indicated by the fact that the intragenic Hind III and Bam HI sites are separated by 340 and 220 bp in the $\beta 4$ genomic and cDNA sequences, respectively. The small knob observed on the 5′ half of some R loops (Figure 6C) is probably due to the presence of the small intervening sequence. Although this intron has not been located precisely, it seems plausible that, as with $\beta 1$ and all other β-globin genes thus far examined (see Discussion), it is located between codons 30 and 31.

Since the mRNA coding sequences of $\beta 3$ and $\beta 4$ share sequence homology with human γ-globin and both genes are expressed in embryonic erythroid cells, it was of interest to compare the sequences

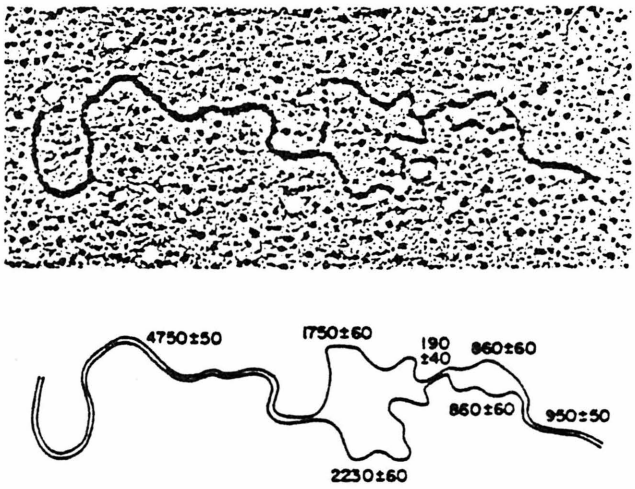

Figure 10. Heteroduplex Analysis of Genes β3 and β4

Eco RI fragments of 2.8 and 3.3 kb containing the 5′ coding and intervening sequence of genes β3 and β4, respectively, were isolated from clone λRβ′G8 and subcloned into pMB9. DNA from these subclones was linearized with Sal I, which cleaves only once within the vector DNA (Rodriguez et al., 1976), mixed, denatured and renatured to form heteroduplexes which were examined in the electron microscope. The results of measuring 22 molecules are given as the average size ± standard deviation in base pairs. The left long duplex arm and right short duplex arm are the pMB9 vector segments. The single-stranded regions begin at the Eco RI sites of the insert. The photograph has the 5′ → 3′ transcriptional orientation of the genes arranged left to right.

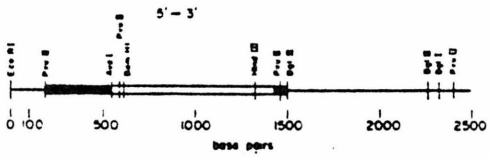

Figure 11. Restriction Map of Gene β2 Showing the Regions Homologous to Cloned β1 cDNA

The map shows the distribution of restriction endonuclease cleavage sites within and flanking gene β2. The black boxes indicate the regions detected by the β1 probe pRcβ1. The white box designates the region of the gene which does not hybridize to either pRcβ1 (β1 cDNA) or to pRB0.65 (β1 intron II) and therefore contains an intervening sequence. The Eco RI site corresponds to the boundary of the 2.13 and 6.3 kb Eco RI fragments shown in Figure 1.

within and flanking the two genes. We therefore examined heteroduplexes between the β3 and β4 genes by electron microscopy. Attempts to analyze heteroduplexes between different bacteriophage λ clones were unsuccessful because of the presence of numerous inverted and tandem repeat sequences (J. Shen and T. Maniatis, manuscript in preparation) which led to the formation of stem-loop structures and tangles of single-stranded DNA. We therefore constructed plasmid subclones containing either the 2.8 or 3.3 Eco RI fragments (Figure 1) which include the 5′ flanking, 5′ coding and intervening sequences of genes β3 and β4, respectively. The plasmid DNAs containing inserts in the same orientation were linearized with Sal I, which cleaves only once at a site within the vector, and were mixed, denatured and reannealed. A typical heteroduplex, shown in Figure 10, reveals only a short region of homology (190 bp) in addition to the pMB9 arms. Based on the measurement of 22 molecules, this homology maps to a region about 860 nt from the intragenic Eco RI site of the two inserts. This distance places the duplex in the coding region between IVS I and IVS II of gene β4 (analogous to exon 2 of β1, Figure 3), assuming that the small IVS in β4 is between codons 30 and 31. Thus the 5′ flanking regions, exon 1 and the large introns of β3 and β4 are not homologous. The heteroduplex data are not sufficiently precise to rule out some homology in the small intervening sequence.

Since the 3′ coding regions of genes β3 and β4 are not contained in any subclones suitable for heteroduplex analysis, we investigated these regions by blot hybridization. By probing appropriate digests of the gene β3 region with either cloned β4 cDNA or genomic DNAs, we were unable to detect hybridization to the 3′ Eco RI fragment of gene β3 (data not shown). The blot hybridization and washing conditions were sufficiently nonstringent (0.75 M Na^+, 68°C) to allow cross-hybridization between the β4 probe and a human ε-globin gene (E. Fritsch, unpublished data) and chick adult and embryonic β-globin DNA (Dodgson, Strommer and Engel, 1979). We conclude that the region of homology between genes β3 and β4 is limited to 190 bp, primarily in exon 2.

Structure of Gene β2

Gene β2 has two striking features. The first is its limited sequence homology to all of the β-like globin probes tested. The second is the apparent absence of a β2 transcript in embryonic or adult erythroid RNA. To determine whether the structure of β2 is similar to that of other β-like globin genes, we constructed a restriction map and localized the regions that hybridize to the adult β-globin cDNA plasmid. The restriction map in Figure 11 represents a 2.5 kb region of DNA containing β2 sequences. We have previously shown that the sequences to the left and right of the Hind III site are homologous to the 5′ and 3′ mRNA coding regions of β1, respectively (Lacy et al., 1979). A 900

bp segment of nonhybridizing DNA lies between the Ava I and the Pvu II sites at 550 and 1460 bp on the map. Since the region homologous to β1 mRNA sequences spans 1300 bp but only 600 bp are needed to encode a typical globin gene, an intervening sequence of at least 700 bp must be present. There is no apparent homology between the large introns of β1 and β2 because no hybridization was detected between a cloned fragment containing the large intervening sequence from β1 and the Ava I/Hind III fragment containing the IVS of β2 (data not shown).

An interesting feature of the map in Figure 11 is the distribution of restriction sites in the intervening sequence region. First, the Eco RI site found at the 3′ junction of the large intron in β1, β3, β4 and many other β-globin genes is missing in β2. Second, the Bam HI site found at codons 98–100 in all mammalian β-globin genes studied thus far lies within the region not homologous to β1 mRNA sequences. This indicates either that the Bam HI site lies in a coding region diverged from that of β1 or that the Bam HI site lies in the large IVS of β2. Both possibilities suggest a structure different from that of other mammalian β-like globin genes.

Discussion

All four isolated rabbit β-like globin genes contain at least one large intervening sequence whose size varies from 573 bp for gene β1 to about 800 bp for genes β3 and β4. A second, small IVS has been identified in β1 and β4. This tripartite structure of the message coding sequence is a common feature of all globin genes investigated, including the mouse β-major and β-minor genes (Konkel et al., 1978; Konkel, Maizel and Leder, 1979); the human δ (Lawn et al., 1978; B. Forget, personal communication), β (Lawn et al., 1978 and our unpublished results), $^{G}\gamma$ and $^{A}\gamma$ genes (Smithies et al., 1978; J. L. Slightom, A. E. Blechl and O. Smithies, personal communication); the adult chicken β-globin gene (Dodgson et al., 1979) and the mouse α-globin gene (Leder et al., 1978). The locations of intervening sequences are identical in every globin gene for which nucleotide sequence data are available, suggesting that acquisition of the two intervening sequences must have been a very early event in the evolution of the gene family, possibly concomitant with the formation of a functional ancestral gene (Leder et al., 1978).

The nucleotide sequence of β1 presented here in conjunction with a sequence independently derived by other investigators (Mantei et al., 1979; Weissmann et al., 1979) provides the first comparison between the complete intron sequences in different alleles of the same gene. It is interesting that there are four nucleotides which differ in the two coding sequences (all of which lead to amino acid replacements) while only two nucleotides are different in the large introns. The possible significance of this observation with regard to intron evolution will be discussed elsewhere. Comparison of a number of mammalian β-like globin genes indicates that the sequences at intron/exon junctions are highly conserved, but the remaining intron sequences in most cases are not. For example, the intervening sequences in all four β-globin genes discussed here share little sequence homology and each intron is present only once in the rabbit genome. Similarly, the large introns in the mouse β^{maj} and β^{min} globin genes do not cross-hybridize (Tiemeier et al., 1978); nor do the large introns of the human δ- and β-globin genes (Lawn et al., 1978). Less information is available regarding sequence divergence of the small introns. While these sequences in the adult rabbit and mouse β-globin genes have diverged considerably (van den Berg et al., 1978), the small introns in the mouse β^{maj} and β^{min} genes are highly conserved (Konkel et al., 1979).

Transcripts from genes β1–β4 are present at vastly different levels during rabbit development. Gene β1 transcripts are found predominantly in adult erythroid tissue, whereas RNA from β3 and β4 is found in embryonic globin-producing cells. Although we have not identified β3 and β4 as embryonic or fetal globin genes, it is clear that β1, β3 and β4 are either differentially transcribed or their transcripts are differentially processed during development. In contrast, no mature mRNA from β2 was detected in anemic adult bone marrow or reticulocyte RNA or in the RNA from 12 day embryos. Thus β2 does not correspond to the rabbit minor adult β-like sequence reported by Clissold et al. (1977), nor is it the analog of the human δ-globin gene which is expressed in immature adult erythroid cells (Wood et al., 1978) or of the sheep β^{c}-globin gene whose expression is induced by anemia (Nienhuis and Benz, 1977).

One possible explanation for the failure to detect mature β2 mRNA transcripts is suggested by the structure of the gene. β2 has a large intervening sequence of 700–900 bp, but the number and location of restriction sites near the junctions between coding and noncoding sequences are different in β2 than in other β-like globin genes. This unusual structure suggests the possibility that β2 may be transcribed into a precursor which is not a suitable substrate for splicing. Experiments to test this interesting possibility are in progress.

The cluster of four rabbit globin genes reported in this paper may not include all of the β-like sequences present in the rabbit genome. When the human γ-globin cDNA plasmid is used as a hybridization probe to Eco RI-digested rabbit DNA in a blot hybridization experiment, three fragments which are not present in the gene cluster are detected (Lacy et al., 1979). The polypeptide information summarized in Table 1 predicts the existence of only three or four β-like globin genes in the rabbit and mouse. The detection of

additional rabbit β-like globin sequences, however, coupled with the isolation of a cluster of four non-adult β-like globin genes from a library of mouse DNA (S. Weaver, N. Haigwood, C. Jahn, C. Hutchison and M. Edgell, personal communication) and a recent report describing the isolation of a mouse fetal hemoglobin (Wu, Sikkema and Zucker, 1978) argue that the rabbit and mouse β-like globin gene families may not be unlike the human globin gene family with regard to gene number and the complexity of developmental regulation.

One interesting question posed by the organization of the rabbit globin gene cluster involves the possible function of the intergenic sequences. Out of a total of 44 kb of chromosomal DNA in the cluster, only 6 kb encode globin genes. The sequences between the genes could encode regulatory signals or nonglobin structural genes. Preliminary experiments indicate the presence of transcripts in total embryonic poly(A) RNA which are homologous to nonglobin sequences throughout the gene cluster, but we have not yet determined whether these homologous sequences are actually transcribed from the rabbit β-globin gene cluster. Characterization of the intergenic regions by cross-hybridization and heteroduplex analysis has revealed a complex array of sequenes which are repeated within the gene cluster (J. Shen and T. Maniatis, manuscript in preparation), some of which hybridize to a nonglobin cDNA clone prepared from embryonic erythroid RNA (E. Butler, unpublished results). The study of transcription and sequence organization in the intergenic regions of the rabbit β-like globin gene cluster may provide insight into the mechanism of differential globin gene expression.

Experimental Procedures

Preparation of RNA from Erythroid Tissues

Anemia was induced in adult male New Zealand white rabbits by phenylhydrazine injection (Crystal, Elson and Anderson, 1974). For RNA isolation from blood or bone marrow, cells were pelleted and washed in 0.14 M NaCl, 0.005 M $MgCl_2$, 0.05 M KCl, 1 U/ml heparin and 0.01 M iodoacetate (pH 7.0). 12 day rabbit embryos were collected from female New Zealand white rabbits following administration of lethal doses of pentobarbital. Total RNA was isolated as described by Ullrich et al. (1977) and modified by J. Hirsch (personal communication). The cells were homogenized in 4 M guanidinium thiocyanate (Fluka/Tridom), 1 M β-mercaptoethanol, 0.1 M sodium acetate (pH 5), 0.01 M EDTA. 1 g CsCl was added per ml and the solution was transferred to a 60Ti polyallomer tube onto a CsCl cushion [one fourth volume CsCl, $\rho = 1.82$, in 0.1 M sodium acetate (pH 5), 0.01 M EDTA]. The RNA was banded by centrifugation in a 60Ti rotor at 45,000 rpm for 28 hr at 20°C. Flakes of RNA which band at about two thirds from the top of the tube were collected, dialyzed against diethylpyrocarbonate-treated 20 mM HEPES (pH 7.5), 1 mM EDTA, 10 mM iodoacetate, phenol-extracted twice in the presence of 0.05% SDS and precipitated with ethanol. Purified embryonic reticulocytes were obtained following the mechanical disaggregation of blood islands, and cytoplasmic RNA was purified by the method of Longacre and Rutter (1977).

RNA Blot Hybridization Experiments

Samples of RNA were separated by size using either one of two procedures. In the first, the RNA was electrophoresed on 1% agarose gels in 0.025 M sodium citrate (pH 3.5), 6 M urea (Lehrach et al., 1977). The RNA was heated at 68°C for about 3 min in 0.025 M citrate, 6 M urea before electrophoresing at 4°C, 150 V, 10.5 hr. To visualize the RNA bands by ethidium bromide staining and to allow transfer of RNA to diazo-cellulose (Alwine et al., 1977), the gel was washed six times in 0.025 M citrate, 0.5 μg/ml ethidium bromide. In the second procedure, RNA was reacted with glyoxal (Matheson, Coleman and Bell, 40% aqueous solution) and run on a 1.5% agarose gel (McMaster and Carmichael, 1977). The RNA from both types of gels was blotted following the procedures described by Wahl, Stern and Stark (1979) except that an arylaminocellulose paper developed by B. Seed was used as the substrate for diazotization. The RNA-cellulose was hybridized with radioactive probes as described (Wahl et al., 1979) with the modifications given by Lacy et al. (1979).

Electron Microscopy

R Loop Analysis

To increase the frequency of full-length molecules in the R loop analysis, DNA from clone λRβ'G8 was cross-linked lightly with psoralen as recommended by Kaback et al. (1979). R loop hybridization reactions (Thomas, White and Davis, 1976) contained 2 ng/μl DNA, 5 ng/μl poly(A) RNA from whole 12 day rabbit embryos (about 5% 9S globin RNA), 70% recrystallized formamide (gift from P. Chandler), 0.5 M NaCl, 0.1 M HEPES (pH 7.5), 0.01 M EDTA. Aliquots of 5 μl were sealed in silanized glass microcapillary tubes, incubated at 59°C for 11–24 hr, added immediately to a 40 μl hyperphase containing 50% formamide, 0.1 M Tris–HCl (pH 8), 0.01 M EDTA, 50 μg/ml cytochrome c, 0.07 ng/μl relaxed circular PM2 marker DNA (a gift from R. Parker) and spread onto a hypophase of 17% formamide, 0.01 M Tris–HCl (pH 8), 0.001 M EDTA (Chow, Kahmann and Kamp, 1977; Tilghman et al., 1978a). The R loop spread was stained and shadowed as described by Davis, Simon and Davidson (1971) and viewed in a Philips EM300 electron microscope. Photographs of molecules were measured with a Hewlett-Packard digitizer; the size of PM2 was taken as 10,040 bp (T. Sargent, personal communication).

Heteroduplex Analysis

DNA from subclones pR2.8 and pR3.3, containing the Eco RI 2.8 and 3.3 kb fragments from genes β3 and β4, respectively, was linearized by digestion with Sal I and extracted with phenol and ether. Samples of DNA (125 ng from each subclone) were mixed, denatured in 0.1 N NaOH, 0.02 M EDTA for 10 min at room temperature, neutralized by adding 0.1 vol 2 M Tris–HCl (pH 7) and renatured after adding an equal volume of formamide (Davis et al., 1971). The final renaturation conditions were 5 ng/μl total DNA, 0.1 M Tris–HCl (pH 8.5), 0.05 M NaCl, 0.01 M EDTA, 50% formamide for 1 hr at room temperature. Cot = 0.054. Heteroduplex DNA (20 ng) was spread as described above for R loops. The photographed structures were measured as the fractional length of the molecule and converted to base pairs using the known sizes of the subclones.

DNA Sequence Analysis

DNA nucleotide sequences were determined by the method of Maxam and Gilbert (1977, 1979). Enzymes used in the analysis were purchased from New England BioLabs, Bethesda Research, PL Biochemicals and Boehringer-Mannheim. The enzyme Hin 389 II was a gift from L. Smith. γ-^{32}P-ATP was purchased from either NEN or ICN.

Acknowledgments

We thank C.-K. J. Shen and B. Seed for advice and helpful discussions, and J. Posakony for advice on DNA sequencing. We would like to acknowledge M. Edgell, O. Smithies and P. Leder for communicating their results before publication. We thank D. Quon for constructing subclones, Cara Berman for technical assistance, S. Van de Woude for assistance with restriction mapping experiments and A. Cortenbach for preparing media and materials.

E.L., R.C.H. and T.M. were supported by an NIH graduate training grant to the California Institute of Technology, a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research and a Rita Allen Foundation Career Development Award, respectively. E.T.B. was supported by an NIH postdoctoral training grant to the

California Institute of Technology. This work was supported by grants from the National Science Foundation to T.M. and A.E., a grant from the American Cancer Society, Massachusetts Division, to A.E. and a grant from the Northwest Area Foundation to T.M.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 14, 1979; revised October 1, 1979

References

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Proc. Nat. Acad. Sci. USA *74*, 5350–5354.

Bastos, R. N. and Aviv, H. (1977). Cell *11*, 641–650.

Berg, D. E., Davies, J., Allet, B. and Rochaix, J. D. (1975). Proc. Nat. Acad. Sci. USA *72*, 3628–3632.

Berk, A. J. and Sharp, P. A. (1977). Cell *12*, 721–732.

Bolivar, F., Rodriguez, R., Greene, P., Betlach, M., Heyneker, H. and Boyer, H. (1977). Gene *2*, 95–113.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). Proc. Nat. Acad. Sci. USA *75*, 4853–4857.

Bricker, J. and Garrick, M. D. (1974). Biochim. Biophys. Acta *351*, 437–441.

Broker, T., Soll, L. and Chow, L. T. (1977). J. Mol. Biol. *113*, 579–589.

Bunn, H. F., Forget, B. G. and Ranney, H. M. (1977). Human Hemoglobins (Philadelphia: W. B. Saunders Co.).

Capp, G. L., Rigas, D. A. and Jones, R. T. (1970). Nature *228*, 278–280.

Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y. and Brownlee, G. G. (1978). Nature *275*, 510–513.

Chow, L. T., Kahmann, R. and Kamp, D. (1977). J. Mol. Biol. *113*, 591–609.

Cissold, P. M., Arnstein, H. R. V. and Chesterton, C. J. (1977). Cell *11*, 353–361.

Crick, F. (1979). Science *204*, 264–271.

Crystal, R. G., Elson, N. A. and Anderson, W. F. (1974). Methods in Enzymology *30*, 101–127.

Curtis, P. J. and Weissmann, C. (1976). J. Mol. Biol. *106*, 1061–1075.

Davis, R. W. (1969). Ph.D. thesis, California Institute of Technology, Pasadena, California.

Davis, R. W., Simon, M. and Davidson, N. (1971). In Methods in Enzymology, *21*, Part D, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 413–428.

Dawid, I. B. and Wahli, W. (1979). Dev. Biol. *69*, 305–328.

Dayhoff, M. O. (1972). Atlas of Protein Sequence and Structure (Washington, D.C.: National Biomedical Research Foundation).

Dodgson, J. B., Strommer, J. and Engel, J. D. (1979). Cell *17*, 879–887.

Efstratiadis, A., Kafatos, F. C. and Maniatis, T. (1977). Cell *10*, 571–585.

Flavell, R. A., Bernards, R., Grosveld, G. C., Hooijmakers-Van Dommelen, H. A. M., Kooter, J. M. and De Boer, E. (1979). In Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, *XIV*, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press), in press.

Forget, B., Cavallesco, J., deRiel, J., Spritz, R., Choudary, P., Wilson, J., Wilson, L., Reddy, V. and Weissman, S. (1979). In Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, *XIV*, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press), in press.

Gale, R., Clegg, J. and Huehns, E. (1979). Nature *280*, 162–164.

Galuzzi, A. (1970). Eur. J. Biochem. *17*, 49–55.

Galuzzi, A. (1971). Nature New Biol. *229*, 142–143.

Gannon, F., O'Hare, K., Perrin, F., LePennec, J. P., Benoist, c., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979). Nature *278*, 428–434.

Garrick, M. D., Hafner, R., Bricker, J. and Garrick, L. (1974). Ann. NY Acad. Sci. *241*, 436–438.

Goldberg, M. (1979). Ph.D. thesis, Stanford University, Stanford, California.

Haynes, J. R., Kalb, F. V., Rosteck, P. and Lingrel, J. B. (1978). FEBS Letters *91*, 173–177.

Huehns, E. R., Dance, N., Beaver, G. H., Keil, J. V., Hecht, F. and Motulsky, A. G. (1964). Nature *201*, 1095–1097.

Jeffreys, A. J. and Flavell, R. A. (1977a). Cell *12*, 429–439.

Jeffreys, A. J. and Flavell, R. A. (1977b). Cell *12*, 1097–1108.

Kaback, D. B., Angerer, L. M. and Davidson, N. (1979). Nucl. Acids Res. *6*, 2499–2517.

Kafatos, F., Efstratiadis, A., Forget, B. and Weissman, S. (1977). Proc. Nat. Acad. Sci. USA *74*, 5618–5622.

Kinniburgh, A. J. and Ross, J. (1979). Cell *17*, 915–921.

Kinniburgh, A. J., Mertz, J. E. and Ross, J. (1978). Cell *14*, 681–693.

Kitchen, H. and Brett, I. (1974). Ann. NY Acad. Sci. *241*, 653–670.

Kleckner, N., Chan, R. K., Tye, B.-K. and Botstein, D. (1975). J. Mol. Biol. *97*, 561–575.

Konkel, D. A., Tilghman, S. M. and Leder, P. (1978). Cell *15*, 1125–1132.

Konkel, D. A., Maizel, J. V., Jr. and Leder, P. (1979). Cell *18*, 865–873.

Kwan, S.-P., Wood, G. and Lingrel, J. (1977). Proc. Nat. Acad. Sci. USA *74*, 178–182.

Lacy, E., Hardison, R. C., Quon, D. and Maniatis, T. (1979). Cell *18*, 1273–1283.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. and Maniatis, T. (1978). Cell *15*, 1157–1174.

Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M. and Leder, P. (1978). Proc. Nat. Acad. Sci. USA *75*, 6187–6191.

Lehrach, H., Diamond, D., Wozney, J. and Boedtker, H. (1977). Biochemistry *16*, 4743–4751.

Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979). Cell *18*, 545–558.

Longacre, S. S. and Rutter, W. J. (1977). J. Biol. Chem. *252*, 2742–2752.

McMaster, G. K. and Carmichael, G. C. (1977). Proc. Nat. Acad. Sci. USA *74*, 4835–4838.

Maniatis, T., Sim, G. K., Efstratiadis, A. and Kafatos, F. C. (1976). Cell *8*, 163–182.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. and Efstratiadis, A. (1978). Cell *15*, 687–701.

Mantei, N., van Ooyen, A., van den Berg, J., Beggs, J., Boll, W., Weaver, R. and Weissmann, C. (1979). In Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, *XIV*, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press), in press.

Marks, P. A. and Rifkind, R. A. (1972). Science *175*, 955–961.

Maxam, A. M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. USA *74*, 560–564.

Maxam, A. and Gilbert, W. (1979). Methods in Enzymology, in press.

Melderis, H., Steinheider, G. and Ostertag, W. (1974). Nature *250*, 774–776.

Nienhuis, A. W. and Benz, E. J. (1977). N. Eng. J. Med. *297*, 1318–

1436.

Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. and Betlach, M. (1976). In Molecular Mechanisms in the Control of Gene Expression, ICN-UCLA Symposium on Molecular and Cellular Biology, 5, D. P. Nierlich, N. J. Rutter and C. F. Fox, eds. (New York: Academic Press), pp. 471–477.

Ross, J. (1976). J. Mol. Biol. *106*, 403–420.

Ross, J. and Knecht, D. A. (1978). J. Mol. Biol. *119*, 1–20.

Shamsuddin, M., Mason, R. G., Cohen, C., Tissot, R. G. and Honig, G. R. (1973). Arch. Biochem. Biophys. *158*, 922–924.

Smith, K. and Lingrel, J. B. (1978). Nucl. Acids Res. 5, 3295–3301.

Smithies, O., Blechl, A. E., Denniston-Thompson, K., Newell, N., Richards, J. E., Slighton, J. L., Tucker, P. W. and Blattner, F. R. (1978). Science *202*, 1284–1289.

Steinheider, G., Melderis, H. and Ostertag, W. (1975). Nature *257*, 714–716.

Thomas, M., White, R. L. and Davis, R. W. (1976). Proc. Nat. Acad. Sci. USA *73*, 2294–2298.

Tiemeier, D. C., Tilghman, S. M., Polsky, F. I., Seidman, J. G., Leder, A., Edgell, M. H. and Leder, P. (1978). Cell *14*, 237–245.

Tilghman, S., Tiemeier, D., Seidman, J., Peterlin, M., Sullivan, M., Maizel, J. and Leder, P. (1978a). Proc. Nat. Acad. Sci. USA *75*, 725–729.

Tilghman, S., Curtis, P., Tiemeier, D., Leder, P. and Weissmann, C. (1978b). Proc. Nat. Acad. Sci. USA *75*, 1309–1313.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. and Goodman, H. (1977). Science *196*, 1313–1319.

van den Berg, J., van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R. and Weissmann, C. (1978). Nature *276*, 37–44.

Wahl, G. M., Stern, M. and Stark, G. R. (1979). Proc. Nat. Acad. Sci. USA *76*, 3683–3687.

Weatherall, D. J. and Clegg, J. B. (1979). Cell *16*, 467–479.

Weissmann, C., Mantei, N., Boll, W., Weaver, R. F., Wilkie, N., Clements, B., Taniguchi, T., van Ooyen, A., van den Berg, J., Fried, M. and Murray, K. (1979). 11th Miami Winter Symposium, in press.

Wold, B., Wigler, M., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Proc. Nat. Acad. Sci. USA, in press.

Wood, K. O. and Lee, J. C. (1976). Nucl. Acids Res. *3*, 1961–1971.

Wood, W. G., Old, J. M., Roberts, A. V. S., Clegg, J. B., Weatherall, D. J. and Quattrin, N. (1978). Cell *15*, 437–446.

Wu, N.-C., Sikkema, D. A. and Zucker, R. M. (1978). Biochim. Biophys. Acta *536*, 306–311.

Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M. and Dunn, A. R. (1979). Cell *16*, 851–861.

Ziff, E. B. and Evans, R. M. (1978). Cell *15*, 1463–1475.

Chapter 4

The Nucleotide Sequence of a Rabbit β Globin Pseudogene

Elizabeth Lacy[+] and Tom Maniatis

Division of Biology, California Institute of Technology,

Pasadena, California 91125

[+]Present address: Department of Zoology, University of Oxford, South Parks Road, Oxford OXI 3PS, United Kingdom.

Running Title: Rabbit β Globin Pseudogene

Summary

We report the nucleotide sequence of a rabbit β globin pseudogene, ψβ2. A comparison of the ψβ2 sequence with that of the rabbit adult β globin gene, β1, reveals the presence of frameshift mutations and premature termination codons in the protein coding sequence which render ψβ2 unable to encode a functional β globin polypeptide. ψβ2 contains two intervening sequences at the same locations in the globin protein coding sequence as β1 and all other sequenced β globin genes. An examination of the DNA sequences at the intron/exon junctions suggests that a putative ψβ2 precursor mRNA could not be normally spliced. We also present a detailed comparison of the flanking and noncoding sequences of ψβ2 and β1 and a discussion of the phylogenetic relationship between these two genes.

Introduction

A cluster of four different β-like globin gene sequences has been isolated from a bacteriophage λ library of rabbit chromosomal DNA in a set of overlapping clones which together contain 44 kilobase pairs (kb) of contiguous DNA (Maniatis et al., 1978; Lacy et al., 1979). The linkage arrangement of the four genes, β1-β4, is shown in Figure 1. Approximately 5-8 kb of DNA separate each gene pair and all four genes are transcribed from the same strand of DNA in the orientation 5'-β4-β3-β2-β1-3' (Lacy et al., 1979).

The nucleotide sequence of gene β1 shows that it encodes the rabbit adult β globin protein (Hardison et al., 1979). The presence of β3 and β4 mature mRNA transcripts in nucleated reticulocytes from the blood islands of 12-day rabbit embryos suggests that these genes encode embryonic and/or fetal β globin polypeptides. Furthermore, the β3 and β4 genes hybridize more efficiently to embryonic than to adult erythroid RNA (Hardison et al., 1979). Therefore, the β1, β3, and β4 genes appear to be differentially expressed during development.

Gene β2 hybridizes more efficiently to adult than to embryonic β globin mRNA sequences. However, no β2 mRNA transcripts have been detected in either adult bone marrow and reticulocytes or in embryonic erythroid cells (Hardison et al., 1979). One possible explanation for the apparent lack of β2 transcripts is that gene β2 is a globin pseudogene, a β-like sequence that does not code for a functional β globin polypeptide. α and β globin genes which cannot be identified with known polypeptides have been discovered in several other mammalian species including mouse (Vanin et al., 1980; Nishioka, Leder and Leder, 1980; Jahn et al., 1980), human (Lauer, Shen and Maniatis, 1980; Proudfoot and Maniatis, accompanying manuscript; Fritsch, Lawn and Maniatis, 1980) and goat (Haynes et al., 1980). To determine whether β2 contains deletions, insertions, or base changes that would generate a nonfunctional globin gene, we have determined its nucleotide sequence and compared

it to that of a known functional β globin gene, β1.

Results and Discussion

The Protein Coding Region

Figure 1 shows the strategy we used to determine the DNA sequence of β2 by the Maxam-Gilbert base-specific chemical degradation technique (Maxam and Gilbert, 1977). An alignment of the nucleotide sequences of β2 and β1 is presented in Figure 2. To make this alignment, deletions and insertions were included in the β2 sequence wherever necessary to maintain identical reading frames and maximum DNA sequence homology in the protein coding regions.

An examination of the two coding sequences reveals that a base has been deleted in β2 at codon 20. This deletion would shift the translational reading frame of a putative β2 mRNA relative to β1 and result in an in-phase terminator, TGA, spanning codons 28 and 29 in β1. Consequently, a β2 mRNA would code for a protein that is only 27 amino acids long. Another terminator occurs in β2 at codon 125 and a second frameshift mutation (a deletion) occurs at codon 128. In addition, base changes have replaced amino acids at several sites conserved in many β globin polypeptides (Dayhoff, 1972). Some of these altered sites are known to function in heme binding and in interactions with α globin chains (Eaton, 1980). Thus, the nucleotide sequence analysis of β2 has revealed the presence of mutations that have rendered it unable to produce a functional β globin polypeptide. We will, therefore, refer to this gene as ψβ2 to indicate that it is a β globin pseudogene.

Comparison of the Noncoding and Flanking Sequences of ψβ2 and β1

There are at least two classes of mutations which could potentially make a gene unable to code for a functional polypeptide. One class consists of insertions/deletions and base changes which result in frameshifts, missense mutations and premature termination codons in the protein coding sequence. As mentioned above, such alterations

are found in the nucleotide sequence of ψβ2. A second class consists of mutations affecting the transcription and processing of nuclear mRNA precursors. In general, genes coding for defective polypeptides may also selectively accumulate mutations preventing the generation and accumulation of a nonfunctional proteins. Sequences which may function in the regulation of transcription and processing have been identified in the noncoding and flanking regions of several eukaryotic genes on the basis of sequence conservation. A comparison of the noncoding and flanking sequences of ψβ2 and β1 is presented in the following sections to determine if ψβ2 has acquired base changes in the sequences conserved in mammalian globin and other eukaryotic genes.

The 5' Flanking Region

Figure 3A shows an alignment of the ψβ2 sequence from base pair -101 to +1 (see Figure 2) with the sequences 5' to the mRNA capping site in the rabbit β1, human β and mouse β^{maj} globin genes. Counting each insertion or deletion as one mismatch, the β1 and ψβ2 5' flanking sequences are 71% homologous. Most of this homology is found in three regions 5' to a putative mRNA capping site in ψβ2: 1) between -12 and +1, 2) between -18 and -35, and 3) between -77 and -84 base pairs. Both regions 2 and 3 are highly conserved among functional adult β globin genes (Figure 3A).

Region 2 contains an A-T rich sequence originally identified in Drosophila histone genes ("the Hogness box") (Goldberg, 1979) and subsequently shown to begin 30 to 31 base pairs 5' to the cap site in most eukaryotic structural genes (see Baker et al., 1979, and Benoist et al., 1980, for references). A comparison of a number of different β-like globin genes has revealed that the A-T rich sequence CATAAA is found in most of these genes, but that the only sequence shared by all the β-like globin genes is PyATAPu. This sequence was therefore designated the ATA box (Efstratiadis et al., 1980). The conserved position of the ATA box relative to the mRNA capping site and its similarity to the "Pribnow box" (TATAAT) of prokaryotic

promoters (Pribnow, 1979) has led a number of investigators to propose that the ATA sequence is involved in the initiation or processing of transcripts from eukaryotic structural genes (see Baker et al., 1979; Grosschedl and Birnstiel, 1980; and Efstratiadis et al., 1980, for a discussion).

Region 3 contains a sequence, CACCCT, which is found in all except one of the adult β globin genes thus far sequenced (Efstratiadis et al., 1980). The fact that the CACCCT sequence is not conserved in the embryonic and fetal human β-like genes (Efstratiadis et al., 1980) nor in other eukaryotic genes (Benoist et al., 1980), including the mouse and human α globin genes (Nishioka and Leder, 1979; Liebhaber et al., 1980) suggests that this sequence is specific to adult β globins.

A region of strong sequence homology found in the 5' flanking region of all mammalian globin genes thus far studied is deleted in ψβ2. With the exception of the human δ globin gene, which has the sequence CCAAC, all the β-like globin genes contain the sequence CCAAT ("CCAAT box") 80 base pairs 5' to the mRNA capping site. In addition, an identical sequence is found at a similar location in both the mouse (Nishioka and Leder, 1979) and human (Liebhaber et al., 1980) α globin genes. The CCAAT box, however, does not appear to be unique to mammalian globin genes. Benoist et al. (1980) have observed a related sequence in a similar location in the 5' flanking sequences of the chicken ovalbumin and conalbumin genes and the adenovirus early IA gene. The alignment in Figure 3A clearly shows that the CCAAT sequence is missing in ψβ2 and that this is the major difference between the 5' flanking region of ψβ2 and that of β1 and other mammalian globin genes.

The 5' Noncoding Region

Figure 3B shows an alignment of the ψβ2 sequence from base pair +1 to +51 (see Figure 2) with the sequences between the mRNA capping site and the initiator ATG in the rabbit β1, human β and mouse β^{maj} globin genes. There are 10 base changes between β1 and ψβ2 in the 5' noncoding region, resulting in an overall homology

of 81%. Thus, the 5' noncoding regions of β1 and ψβ2 are as conserved as those of the mouse adult β^{maj} and β^{min} genes, which have a homology of 80% (Konkel, Maizel and Leder, 1979), but less conserved than the 5' noncoding regions of the two human adult genes, β and δ, which share a homology of 92% (Efstratiadis et al., 1980).

The sequence comparisons presented in Figures 3A and 3B identify a putative mRNA capping site in ψβ2 and suggest that the first three nucleotides of a mature β2 mRNA would be AUG. A second AUG is found in the putative 5' noncoding region of ψβ2 at the site expected for the initiation of protein synthesis. Since translation initiates at the AUG closest to the 5' end of a mRNA (Kozak, 1978), it is conceivable that a ψβ2 message beginning with a capped AUG might not be translated correctly or efficiently.

The sequence CUUPyUG, first noted by Baralle and Brownlee (1978), is found in the 5' noncoding region of all mammalian α and β globin genes for which sequence information is available (Efstratiadis et al., 1980). Since the CUUPyUG sequence shares homology with a conserved purine-rich region at the 3' end of eukaryotic 18S rRNAs, Hagenbüchle et al. (1978) have speculated that this sequence may function in the initiation of translation. ψβ2 contains a CUUPyUG sequence (UAUUUG) which differs in two bases from the sequence in β1 (CUUUUG) (Figure 3B). An alignment of the β1 and ψβ2 sequences with the purine-rich sequence in 18S rRNAs - 3' UAGGAAGGCGU 5' - (Hagenbuchle et al., 1978) indicates that the ψβ2 sequence would form a less stable hybrid than the β1 sequence.

Between the putative capping site and initiator ATG, ψβ2 contains 48 base pairs (bp) whereas β1 contains 53 bp (see Figure 3B). The difference in length can be accounted for by a deletion of 5 base pairs from β1, 3 base pairs before the ATG. An examination of the sequence in this region reveals that part of a direct repeat - CAGACAGA - has been lost in ψβ2. An analysis of the DNA sequence at hot spots

for spontaneous mutations in the lac I gene of E. coli suggests that direct repeats may be involved in the generation of deletions (Farabaugh and Miller, 1978). A similar loss of direct repeats in noncoding sequences has been noted by Efstratiadis et al. (1980) when making pairwise comparisons of the rabbit, human, and mouse β-like globin genes. To obtain an alignment that maximizes homology in any one gene pair, it was often necessary to assume that deletions had occurred in one of the two genes. They observed that deletion sites are flanked by short direct repeats and that a deletion removes one repeat completely and part or none of the other repeat. A mechanism, which is based on models proposed for prokaryotic systems (Farabaugh and Miller, 1978; Streisinger et al., 1966), is presented by Efstratiadis et al. (1980) to explain how deletions might be generated from the mispairing of repeats during replication.

Intervening Sequences

The mRNA coding region in all α and β globin genes studied thus far is interrupted by two intervening sequences (Tilghman et al., 1978a; Jeffreys and Flavell, 1977; Lawn et al., 1978; Konkel, Tilghman and Leder, 1978; Konkel et al., 1979; Smithies et al., 1978; Mantei et al., 1979; Hardison et al., 1979; Proudfoot and Baralle, 1979; Dodgson, Strommer and Engel, 1979; Efstratiadis et al., 1980; Leder et al., 1978; Lauer et al., 1980). DNA sequence analyses of β globin genes have shown that in all cases one intron (IVS 1) is located between codons for amino acids 30 and 31, while the second, larger intron (IVS 2) is located between codons 104 and 105. Therefore, a globin mRNA sequence is encoded in the genome in three discontinuous blocks: exons I, II, and III. We previously identified an intron of 700-900 bp in the ψβ2 gene by hybridization and restriction mapping experiments (Hardison et al., 1979). The existence of this large intron (IVS 2) is confirmed by comparing the nucleotide sequences of the ψβ2 and β1 genes (Figure 2). This comparison also shows that ψβ2 contains a second, smaller intervening sequence (IVS 1) and that the locations of both introns

in ψβ2 are identical to those found in all other β-like globin genes.

IVS 1 of ψβ2 is 26 bp smaller than IVS 1 of β1 (100 vs. 126 bp). Most of this difference in size can be accounted for by assuming that a 21 bp region immediately adjacent to the 3' end of IVS 1 was deleted in ψβ2 (Figure 2). We note that the pentanucleotide GGCTG occurs near both end points of the putative deletion. The previous section on the 5' noncoding region of ψβ2 discussed the possible involvement of short, direct repeats in the generation of deletions (Farabaugh et al., 1978; Marotta et al., 1977; Efstratiadis et al., 1980). It is possible that this deletion was produced by the proposed mispairing mechanism.

IVS 2 of ψβ2 is approximately 200 bp larger than IVS 2 of β1. Only a portion (457 bp) of the second intron in ψβ2 has been sequenced, including 176 bp from the 5' junction and 281 bp from the 3' junction. Previously, we reported that the large intron of the ψβ2 gene does not hybridize to the β1 gene (Hardison et al., 1979). This observation is consistent with the fact that, with the exception of a few nucleotides at the intron/exon junctions, we were unable to make any reasonable alignments between IVS 2 of β1 and the available nucleotide sequence of IVS 2 in ψβ2.

A consensus sequence has been derived from an analysis of the DNA sequences at the intron/exon junctions of several eukaryotic genes (Lerner et al., 1980; Crick, 1979). The consensus sequence defines a common splicing frame which predicts, with three exceptions, that an intron begins with GT and ends with AG (Breathnach et al., 1978; Lerner et al., 1980). An examination of the four intron/exon junctions in ψβ2 indicates that, if this pseudogene is transcribed, the introns could not be excised according to the splicing frames specified by the consensus sequence in the β1 gene.

Figure 4 presents a comparison of the DNA sequences at the intron/exon boundaries of ψβ2 and β1. The 5' junction of IVS 1 in ψβ2 differs in 4 out of 15 bp from the β1 junction. The presence of the dinucleotide GT in ψβ2 suggests that

the 5' junction could serve as a substrate for splicing. However, the occurrence of a deletion at the 3' end of IVS 1 makes it unlikely that a ψβ2 transcript could be normally spliced. An AG dinucleotide is present near the IVS 1/exon II junction, but its location within exon II would lead to a splice which would alter the normal globin translational reading frame.

The sequence at the 5' junction of IVS 2 in ψβ2 shares 13 out of 15 bp with the β1 sequence and is consistent with a GT/AG splicing frame. The 3' junction, on the other hand, can best be aligned with the β1 sequence if a deletion is included in ψβ2 at the G in the AG dinucleotide. However, if the splice is placed after the AG in the second codon of exon III, the ψβ2 sequence lines up with the consensus sequence in Figure 4 in 9 out of 11 bases. This splice would not disrupt the translational reading frame, but it would delete the codons for amino acids 105 and 106 in the resulting mRNA.

The mouse, rabbit, and human β globin genes are transcribed into 15-17S nuclear RNA precursors which contain both intron sequences. Subsequently, the intervening sequences are spliced out in steps to generate mature mRNAs (Tilghman et al., 1978b; Kinniburg, Mertz and Ross, 1978; Kinniburg and Ross, 1979; Flavell et al., 1979; Hardison et al., 1979; Maquat et al., 1980; Kantor, Turner and Nienhuis, 1980). The observation that a ψβ2 transcript may not be a suitable substrate for RNA processing raises the possibility that ψβ2 transcripts may accumulate in the nucleus as unspliced precursor mRNAs. Kantor et al. (1980) and Maquat et al. (1980) have reported that the deficiency of mature β globin mRNAs in several β^+ thalassemia patients results from the inefficient processing and consequent accumulation of nuclear precursors. However, the nucleotide sequences have not been determined for the β^+ thalassemia genes in these studies and thus, it is not known whether there are base changes in the intron/exon junctions.

ψβ2 mature mRNAs are not detected in anemic adult bone marrow and reticulocyte RNA or in RNA from 12-day embryos (Hardison et al., 1979). To determine whether ψβ2 precursor mRNAs are present in rabbit erythroblasts, a blot of total and poly(A) RNA (Alwine, Kemp and Stark, 1977) from the bone marrow of an anemic rabbit was hybridized to the second intron from both ψβ2 and β1. β1, but not ψβ2, precursor RNA molecules were detected (EL, unpublished results). This observation indicated that, if ψβ2 is transcribed, its mRNA precursors are present at less than a tenth the concentration of β1 precursors in anemic adult rabbit bone marrow cells. The lack of ψβ2 cytoplasmic and nuclear precursor mRNA sequences in embryonic and adult erythroid cells (Hardison et al., 1979), suggests that ψβ2 may not be transcribed in vivo or, if it is, its transcripts are rapidly degraded.

The 3' Noncoding Region

A termination codon, TAA, is found in ψβ2 at the end of the putative protein coding sequence (Figure 2). A different termination codon, TGA, is used in the β1 sequence. Figure 2 shows a comparison of the noncoding sequences 3' to the termination codon in ψβ2 and β1. The two sequences clearly do not share extensive homology in this region, since an alignment requires the introduction of several deletions or insertions. The only sequence common to ψβ2 and β1 is the hexanucleotide AATAAA, which precedes the poly(A) addition site in several eukaryotic mRNAs (Proudfoot and Brownlee, 1976).

This lack of significant sequence homology is consistent with the results of blot hybridization experiments which indicate that the 3' end of ψβ2 cannot be detected with a labeled β1 probe (Lacy et al., 1979). It should be noted, however, that the divergence between the 3' untranslated regions of ψβ2 and β1 is not necessarily a direct consequence of ψβ2 evolving as a pseudogene. The 3' noncoding regions of the human, mouse, and rabbit β-like globin genes sustained many deletions and insertions during their evolution (Efstratiadis et al., 1980).

Phylogenetic Relationships

To learn more about the evolutionary relationship between ψβ2 and β1, we have calculated the divergence between the protein coding sequences of these two genes. Over evolutionary time, the DNA and protein sequences of homologous genes accumulate changes and diverge. In general, two types of base changes occur: those that generate amino acid replacements (replacement substitutions) and those that produce synonymous codons (silent substitutions). The accumulation of replacement substitutions in a particular family of proteins is proportional to time, although different proteins accumulate replacements at different rates. This property, referred to as the evolutionary clock hypothesis, was first established by comparing the amino acid sequences of related proteins (for a review see Wilson, Carlson and White, 1977). Recently, Perler et al. (1980) have developed an improved method for calculating the divergence between the protein coding sequences of homologous genes. Their analysis of the divergence of insulin and globin gene sequences also indicates that replacement substitutions provide a reliable evolutionary clock and that the rate of accumulation of silent substitutions, which is greater than that of replacements, cannot serve as a clock over a long time scale. However, silent substitutions may be able to serve as a clock when the sequences in question have diverged in recent evolutionary time (i.e., within the last 85 million years) (Perler et al., 1980).

To confirm that ψβ2 is more closely related to the adult gene, β1, than to the embryonic genes, β3 and β4, we have estimated the amount of divergence between the protein coding sequences of the four rabbit β-like globin genes (Table 1, Part A). Our analysis includes only the second exons because the available nucleotide sequence for genes β3 and β4 (provided by R. Hardison and E. Butler, unpublished results) does not include the first and third exons. The percent divergences were calculated from pairwise comparisons following the procedures described in Perler et al. (1980). The resulting percentages were corrected for multiple events, assuming

that base changes are Poisson-distributed (Salser, 1977; Perler et al., 1980). It is clear from the results for replacement site substitutions listed in Part A of Table 1 that ψβ2 shares more homology with β1 than with either β3 or β4.

The methodology of Perler et al. (1980) has been used to calibrate an evolutionary clock for globins (Efstratiadis et al., 1980). According to this clock, a 1% change in replacement sites requires 10 million years to become fixed in two initially identical genes, an UEP (unit evolutionary period) of 10. Therefore, the approximate time of divergence for two β globin genes can be estimated from the percent divergence calculated from replacement site substitutions. A percent divergence value for a pair of genes actually represents the sum of the sequence differences each gene has accumulated as it diverged from the common ancestral gene. In addition, the time of divergence predicted for a pair of genes does not necessarily correspond to the time when the gene duplication event took place. If the initial products of the gene duplication were corrected against each other for an unknown period of time, the time of divergence would correspond to the time of the last gene correction event (Efstratiadis et al., 1980).

Part B of Table 1 compares the percent divergence calculated from an alignment of the entire protein coding sequences of ψβ2 and β1 with the percent divergence calculated for the human δ and β globin genes (Efstratiadis et al., 1980). ψβ2 and β1 exhibit a greater degree of replacement site divergence (16.9%) than do δ and β (3.7%) although both pairs of genes have accumulated a similar number of silent substitutions. The percent divergence in silent substitutions for the δ/β gene pair predicts the same time of divergence as the replacement site substitutions, 40 MY (million years) ago (Efstratiadis et al., 1980). According to the clock constructed from silent substitutions, β1 and ψβ2 also began to diverge quite recently, i.e., 55 MY ago. In contrast, the replacement substitutions predict that the rabbit adult β globin genes diverged approximatley 170 MY ago. One interpretation of the

discrepancy in the predicted times of divergence is that the replacement sites in $\psi\beta 2$ were not under selective pressure for the entire period of time since the $\beta 1$-$\psi\beta 2$ divergence. In this case, the percent divergence calculated from silent substitutions would more accurately predict the time of divergence of $\beta 1$ and $\psi\beta 2$.

If $\psi\beta 2$ was a pseudogene from the time it started to diverge from $\beta 1$ 55 MY ago, we might expect replacement sites to be equivalent to silent sites. In other words, all base changes would be silent substitutions and replacement substitutions would accumulate at a similar rate to silent substitutions. Since the percent divergence in replacement sites is approximately two-fold lower than in silent sites, $\psi\beta 2$ probably diverged for a time from $\beta 1$ as a functional gene before acquiring mutations that rendered it nonfunctional and unselected.

It is possible to estimate the time at which $\psi\beta 2$ became a pseudogene if we consider the divergence between two related genes as the sum of the divergences of each gene from the common ancestor. The rate of change in replacement sites between two selected globin genes is 0.1%/MY, while the rate of change in silent sites is 0.8%/MY. Thus, the 16.9% divergence between $\beta 1$ and $\psi\beta 2$ would be the sum of the percent divergence accumulated by $\beta 1$ as a selected gene for the last 55 MY (55 x 0.05) and of the percent divergence accumulated by $\psi\beta 2$ during N years under selection (N x 0.05) and during 55-N years as an unselected globin gene (55-N x 0.4). This calculation predicts that $\psi\beta 2$ diverged as a functional globin for 22 MY. Then approximately 33 MY ago, well after the mammalian radiation (85 MY ago), $\psi\beta 2$ began to diverge as an unselected globin gene.

Concluding Remarks

The occurrence of pseudogenes in eukaryotic gene clusters was first reported in the oocyte 5S DNA repeat unit of Xenopus laevis (Jacq, Miller and Brownlee, 1977; Miller et al., 1978). In each repeat unit there are two 5S gene sequences. One gene encodes the complete oocyte 5S RNA while the second gene (the pseudogene) differs

from the first by 10 base substitutions and a 3' terminal deletion of 19 bp. Since pseudogene transcripts were not detected in vivo, the presence of a pseudogene in the 5S repeat unit may have no functional significance. Rather, its presence may simply be a consequence of the mechanisms of duplication and gene correction which are thought to act on multigene families (Miller et al., 1978).

The aberrant nucleotide sequence of ψβ2 and the apparent absence of ψβ2 transcripts in vivo indicate that this gene is also a pseudogene. Similar pseudogenes have been identified in other mammalian globin gene clusters. Nucleotide sequence analyses of a human α globin pseudogene, ψα1 (Lauer et al., 1980; Proudfoot and Maniatis, accompanying manuscript), and a mouse β globin pseudogene, waw-a (Jahn et al., 1980), revealed the presence of deletions, insertions and base changes which altered the translational reading frame of the protein coding sequences. ψα1 contains two intervening sequences at positions identical to those of functional α globin genes. However, as in the case of ψβ2, splicing cannot occur at the intron/exon junctions of ψα1 in accordance with the splicing frames specified by the consensus sequence of Lerner et al. (1980). The DNA sequence of another α pseudogene in mouse (designated α30.5 by Vanin et al., 1980, and α-3 by Nishioka et al., 1980), demonstrated not only the presence of premature termination codons and frameshift mutations in the protein coding sequence but also the precise excision of both intervening sequences.

The location of the mouse α globin pseudogene with respect to the functional mouse α globin genes is unknown. However, the locations of ψβ2, waw-a, and ψα1 in their respective globin gene clusters have been established. In each case, the pseudogene is found between the embryonic (or fetal) genes and the adult genes. A β-like sequence, ψβ1, which cannot be identified with a known globin polypeptide, is found in the human β-like globin gene cluster, also between the adult and fetal genes (Fritsch et al., 1980). The equivalent position occupied by pseudogenes in the different mammalian globin gene linkage groups suggests that pseudogenes may

have some, as yet unidentified, function in the gene clusters. However, the observation that $\psi\beta 2$ and $\beta 1$ diverged well after the time at which adult and embryonic (fetal) specific globin sequences began to appear (approximately 200 MY ago [Efstratiadis et al., 1980]) indicates that the creation of the rabbit β pseudogene was not coincident with the formation of the gene cluster. In addition, the fact that the rabbit $\psi\beta 2$ gene and the human α globin gene, $\psi\alpha 1$, diverged relatively recently from their functional counterparts suggests that pseudogenes independently arose in different gene clusters. Consequently, the pseudogenes that have been identified in different mammalian globin gene families cannot resemble each other in location or function as a reuslt of a common evolutionary history.

As in the case of the 5S pseudogene, globin pseudogenes may have no function but may simply be products of gene duplication and subsequent sequence divergence. A common feature of globin gene clusters, which reflects their evolution by gene duplication, is the occurrence of two adjacent genes which are coordinately expressed during a given developmental stage. In humans, the fetal β-like and the adult α-like globin polypeptides are encoded in pairs of nearly identical and closely linked genes, $^{G}\gamma\text{-}^{A}\gamma$ and $\alpha 1\text{-}\alpha 2$, respectively (Slightom, Blechl and Smithies, 1980; Lauer et al., 1980). Both members of the α globin gene pair are expressed at similar levels, whereas the ratio of $^{G}\gamma$ to $^{A}\gamma$ expression varies during the fetal to adult switch (Comi et al., 1980). In contrast, the similar but non-identical adult β^{maj} and β^{min} globin genes of the mouse are expressed at quite different levels in some strains of mice (Russell and McFarland, 1974). In other strains, only one adult β globin polypeptide is thought to be expressed, although two genes can be detected by hybridization. One possible explanation of this observation is that one of the two adult globin genes is defective (Weaver et al., 1979). A similar situation is observed in the human $\delta\text{-}\beta$ globin gene pair. The δ protein comprises less than 2.5% of the adult β globin polypeptides in erythrocytes (Bunn et al., 1977). Although the δ globin gene is found in many

primate species, this gene appears to be silent in certain Old World monkeys (Martin et al., 1980). Thus, there is a wide spectrum of divergences in the structure and levels of expression within different globin gene pairs. Perhaps globin pseudogenes represent the most extreme case of divergence of structure and function. In fact, there may be other pseudogenes in globin gene clusters which have diverged so extensively that they no longer are detected by globin gene hybridization probes.

In summary, the structural analysis of a number of different globin gene clusters suggests that globin gene families are in evolutionary flux (see Lauer et al., 1980, for discussion). Perhaps pseudogenes are simply a natural consequence of the mechanisms by which multigene families evolve.

Experimental Procedures

The procedures employed to construct plasmid subclones, to prepare plasmid DNAs, and to derive restriction endonuclease cleavage maps have been described elsewhere (Lacy et al., 1979; Lawn et al., 1978). DNA fragments were sequenced using the procedure of Maxam and Gilbert (1977). Fragments were labeled with ^{32}P at their 5' ends with T4 polynucleotide kinase (Boehringer-Mannheim) following dephosphorylation with bacterial alkaline phosphatase (Bethesda-Research). Fragments were labeled at their 3' ends with the Klenow fragment (Klenow and Henningston, 1970) of E. coli DNA polymerase 1. From 1 to 20 μg of DNA were labeled with 50 μCi each of two $\alpha^{32}P$-dXTPs (2000-3000 Ci/mmole, Amersham) in the presence of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 2 units of Klenow polymerase (Boehringer-Mannheim) for 15 min at 25°C. The reaction was chased for 10 min at 25°C with 0.1 mM of all four unlabeled deoxynucleoside triphosphates and an additional unit of enzyme when labeling fragments with 3' recessive ends.

Acknowledgements

We thank B. Seed, P. F. R. Little and C.-K. J. Shen for helpful discussions and J. Posakony for the computer analysis of the replacement site and silent site substitutions. E.L. and T.M. were supported by a NIH graduate training grant to the California Institute of Technology and the Rita Allen Foundation, respectively. This work was supported by a grant from the National Science Foundation.

References

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA 74, 5350-5354.

Baker, C. C., Herisse, J., Courtois, G., Galibert, F., and Ziff, E. (1979). Messenger RNA for the Ad 2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end. Cell 18, 569-580.

Baralle, F. E., and Brownlee, G. G. (1978). AUG is the only recognizable signal sequence in the 5' non-coding regions of eukaryotic mRNA. Nature 274, 84-87.

Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980). The ovalbumin gene - sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Nat. Acad. Sci. USA 75, 4853-4857.

Bunn, H. F., Forget, B. G., and Ranney, H. M. (1977). Human Hemoglobins (Philadelphia: W. B. Saunders Co.).

Comi, P., Giglioni, B., Ottolenghi, S., Gianni, A. M., Polli, E., Barba, P., Covelli, A., Migliaccio, G., Condorelli, M., and Peschle, C. (1980). Globin chain synthesis in single erythroid bursts from cord blood: studies on $\gamma \rightarrow \beta$ and $^{G}\gamma \rightarrow {}^{A}\gamma$ switches. Proc. Nat. Acad. Sci. USA 77, 362-365.

Crick, F. (1979). Split genes and RNA splicing. Science 204, 264-271.

Dayhoff, M. O. (1972). Atlas of Protein Sequence and Structure (Washington, D.C.: National Biomedical Research Foundation).

Dodgson, J. B., Strommer, J., and Engel, J. D. (1979). Isolation of the chicken β-globin gene and a linked embryonic β-like globin gene from a chicken DNA recombinant library. Cell 17, 879-887.

Eaton, W. A. (1980). The relationship between coding sequences and function in hemoglobin. Nature 284, 183-185.

Efstratiadis, A., et al. (1980). DNA sequence comparison of the human β-like globin genes. Cell, submitted.

Farabaugh, P. J., and Miller, J. H. (1978). Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lac I gene of Escherichia coli. J. Mol. Biol. 126, 847-863.

Flavell, R. A., Bernards, R., Grosveld, G. C., Hoeijmakers-VanDommelen, H. A. M., Kooter, J. M., De Boer, E., and Little, P. F. R. (1979). The structure and expression of globin genes in rabbit and man. In Eukaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, XIV, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press) pp. 335-354.

Fritsch, E. F., Lawn, R. M., and Maniatis, T. (1980). Molecular cloning and characterization of the human β-like globin gene cluster. Cell 19, 959-972.

Goldberg, M. (1979). Ph.D. Thesis, Stanford University.

Grosschedl, R., and Birnstiel, M. L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Nat. Acad. Sci. USA 77, 1432-1436.

Hagenbüchle, O., Santer, M., Steitz, J. A., and Mans, R. J. (1978). Conservation of the primary structure at the 3' end of 18S rRNA from eucaryotic cells. Cell 13, 551-563.

Hardison, R. C., Butler, E. T., Lacy, E., Maniatis, T., Rosenthal, N., and Efstratiadis, A. (1979). The structure and transcription of four linked rabbit β-like globin genes. Cell 18, 1285-1297.

Haynes, J. R., Smith, K., Rosteck, P., Schon, E. A., Gallagher, P. M., Burks, D. J., and Lingrel, J. B. (1980). The isolation of the β^A, β^C, and γ globin genes and a presumptive embryonic globin gene from a goat DNA recombinant library. J. Biol. Chem., in press.

Jacq, C., Miller, J. R., and Brownlee, G. G. (1977). A pseudogene structure in 5S DNA of Xenopus laevis. Cell 12, 109-120.

Jahn, C. L., Hutchinson, C. A. III, Phillips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F., and Edgell, M. H. (1980). DNA sequence organization of the β-globin complex in the BALB/c mouse. Submitted for publication.

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.

Kantor, J. A., Turner, P. H., and Nienhuis, A. W. (1980). $Beta^+$ thalassemia: mutations which affect processing of the β-globin mRNA precursor. Cell, submitted.

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

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 large intervening sequence. Cell 17, 915-921.

Klenow, H., and Henningston, I. (1970). Selective elimination of the exonuclease activity of the DNA polymerase from E. coli B by limited proteolysis. Proc. Nat. Acad. Sci. USA 65, 168-175.

Konkel, D. A., Tilghman, S. M., and Leder, P. (1978). The sequence of the chromosomal mouse β-globin major gene: homologies in capping, splicing and poly(A) sites. Cell 15, 1125-1132.

Konkel, D. A., Maizel, J. V., and Leder, P. (1979). The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18, 865-873.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 15, 1109-1123.

Lacy, E., Hardison, R. C., Quon, D., and Maniatis, T. (1979). The linkage arrangement of four rabbit β-like globin genes. Cell 18, 1273-1283.

Lauer, J., Shen, C.-K. J., and Maniatis, T. (1980). The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. Cell 20, 119-130.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T. (1978). The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell 15, 1157-1174.

Lawn, R. M., Efstratiadis, A., O'Connell, C., and Maniatis, T. (1980). The nucleotide sequence of the human β-globin gene. Cell, submitted.

Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M., and Leder, P. (1978). Comparison of cloned mouse α- and β-globin genes: conservation of intervening sequence locations and extragenic homology. Proc. Nat. Acad. Sci. USA 75, 6187-6191.

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

Liebhaber, S., Goosens, M., and Kan, Y. W. (1980), submitted for publication.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687-701.

Mantei, N., van Ooyen, A., van den Berg, J., Beggs, J. D., Boll, W., Weaver, R. F., and Weissmann, C. (1979). Synthesis of rabbit β-globin-specific RNA in mouse L cells and yeast transformed with cloned rabbit chromosomal β-globin DNA. In Eucaryotic Gene Regulation, ICN-UCLA Symposium on Molecular and Cellular Biology, XIV, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press) pp. 477-498.

Maquat, L. E., Kinniburgh, A. J., Beach, L. R., Honig, G. R., Lazerson, J., Ershler, W. B., and Ross, J. (1980). Processing of the human β-globin mRNA precursor to mRNA is defective in three patients with β^+ thalassemia. Proc. Nat. Acad. Sci. USA, submitted for publication.

Marotta, C. A., Wilson, J. T., Forget, B. G., and Weissman, S. M. (1977). Human β-globin messenger RNA. III. Nucleotide sequences derived from complementary DNA. J. Biol. Chem. 252, 5040-5053.

Martin, S. L., Zimmer, E. A., Kan, Y. W., and Wilson, A. C. (1980). Silent δ globin gene in Old World monkeys. Proc. Nat. Acad. Sci. USA 77, in press.

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA 74, 560-564.

Miller, J. R., Cartwright, E. M., Brownlee, G. G., Federoff, N. V., and Brown, D. D. (1978). The nucleotide sequence of oocyte 5S DNA in Xenopus laevis. II. The GC-rich region. Cell 13, 717-725.

Nishioka, Y., and Leder, P. (1979). The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell 18, 875-882.

Nishioka, Y., Leder, A., and Leder, P. (1980). An unusual alpha globin-like gene that has cleanly lost both globin intervening sequences. Proc. Nat. Acad. Sci. USA, in press.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980). The evolution of genes: the chicken preproinsulin gene. Cell, in press.

Pribnow, D. (1979). Genetic control signals in DNA. In Biological Regulation and Development, 1, R. Goldberger, ed. (New York: Plenum Press) pp. 219-277.

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

Proudfoot, N., and Baralle, F. (1979). Molecular cloning of the human ε-globin gene. Proc. Nat. Acad. Sci. USA 76, 5435-5439.

Proudfoot, N. J., and Maniatis, T. (1980). The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. Cell, submitted.

Russell, E. S., and McFarland, E. C. (1974). Genetics of mouse hemoglobins. Ann. N.Y. Acad. Sci. 244, 25-38.

Salser, W. (1977). Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. 42, 985-1002.

Slightom, J., Blechl, A., and Smithies, O. (1980). Human fetal $^{G}\gamma$ and $^{A}\gamma$ globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell, in press.

Smithies, O., Blechl, A., Denniston-Thompson, K., Newell, N., Richards, J., Slightom, J., Tucker, D., and Blattner, F. (1978). Cloning human fetal γ-globin and mouse α-type globin DNA: characterization and partial sequencing. Science 202, 1284-1289.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., Inouye, M. (1966). Framshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sillivan, M., Maizel, J., and Leder, P. (1978a). Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc. Nat. Acad. Sci. USA 75, 725-729.

Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P., and Weissman, C. (1978b). The intervening sequence of a mouse β globin gene is transcribed within the 15S β-globin mRNA precursor. Proc. Nat. Acad. Sci. USA 75, 1309-1313.

Vanin, E. F., Goldberg, G. I., Tucker, P. W., and Smithies, O. (1980). A mouse alpha globin-related pseudogene (ψα30.5) lacking intervening sequences. Nature, in press.

Weaver, S., Haigwood, N. L., Hutchison, C. A. III, and Edgell, M. H. (1979). DNA fragments of the Mus musculus β globin haplotypes Hbb^s and Hbb^d. Proc. Nat. Acad. Sci. USA 76, 1385-1389.

Wilson, A. C., Carlson, S. S., and White, T. J. (1977). Biochemical evolution. Ann. Rev. Biochem. 46, 573-639.

Table 1. Corrected Percent Divergences of the Coding Sequences of the Rabbit β-like Globin Genes

A. Exon II

Gene Pair	Replacement Sites	Silent Sites
ψβ2/β1	14.8	33.3
ψβ2/β3	29	79.3
ψβ2/β4	26.9	74.1

B. Total Coding Region

Gene Pair	Replacement Sites	Silent Sites
ψβ2/β1	16.9	44.3
Human δ/β	3.7	31.8

Legend to Table 1

The corrected percent divergence between each pair of coding sequences was calculated as described by Perler et al. (1980) except that the codons UUA (Leu), UCG (Ser) and AGA (Arg) were included in the calculation. The sequence data for genes β1, β3 and β4 are from Hardison et al. (1979), R. Hardison (unpublished results) and E. Butler (unpublished results), respectively. The corrected percent divergence for the human δ and β gene pair is from Efstratiadis et al. (1980). The calculations in Part A only include exon II of the four rabbit β-like globin genes.

Figure 1. Linkage Arrangement of the Rabbit β-like Globin Genes and the Strategy for Determining the Nucleotide Sequence of ψβ2

The top line shows the linkage arrangement of the four rabbit β-like globin genes. The direction of transcription of all four genes is 5' → 3' from left to right. Gene ψβ2 is contained within a 2.24 kb fragment generated by a limit Eco RI and partial Bgl II digestion. A plasmid subclone of this fragment (pRI·Bgl II 2.24) was used in the sequence analysis. A fine structure restriction enzyme map of this region is shown in the bottom half of the figure. Only those restriction enzyme sites used in deriving the sequence are indicated. Base pairs (bp) are numbered in both directions from the putative mRNA capping site (0 bp). The putative mRNA coding region of the gene (filled boxes) and intervening sequences (open boxes) are indicated. The open box bordered by a dotted line denotes the sequenced portion of the putative 3' untranslated region. The amino acid codon numbers were assigned on the basis of a DNA sequence comparison between the β1 and ψβ2 genes. The horizontal arrows below the map denote the regions of the DNA that were sequenced. The arrows are labeled with a 3' or 5' to indicate whether the restriction fragments were radioactively labeled at their 5' or 3' ends.

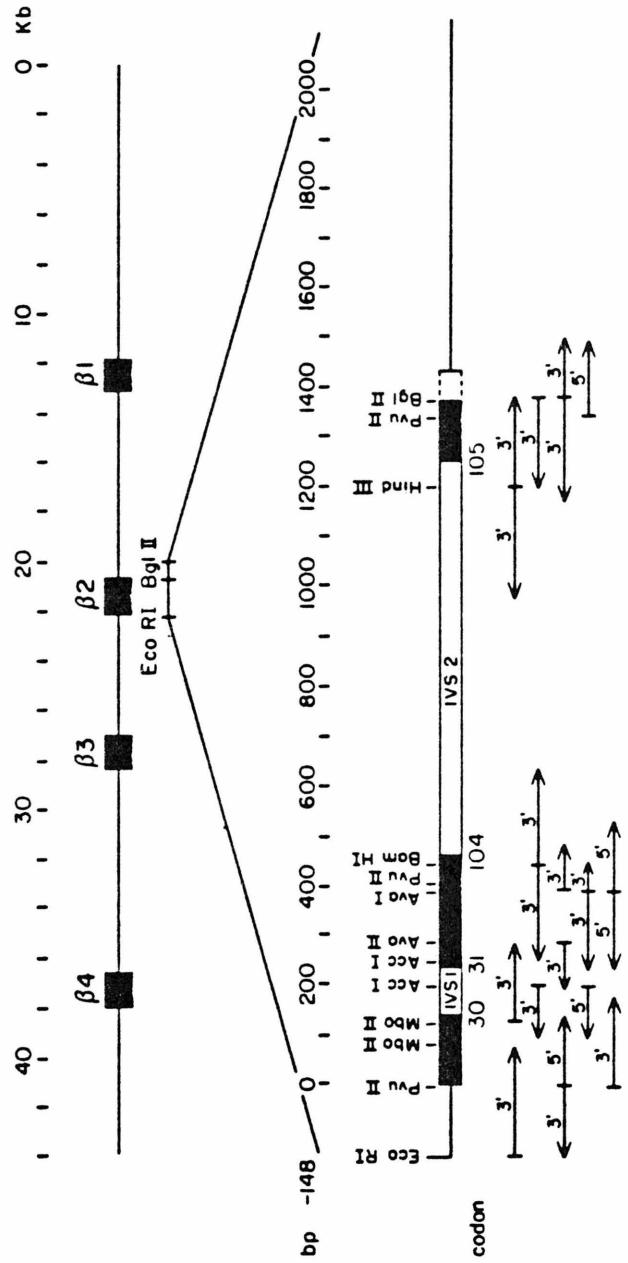
0 Kb
10
20
30
40
β1
β2
β3
β4
Eco RI
Bgl II
0
200
400
600
800
1000
1200
1400
1600
1800
2000
bp -148
Eco RI
Pvu II
Mbo II
Mbo II
Acc I
Acc I
Ava II
Ava I
Pvu II
Bam HI
IVS1
IVS 2
Hind III
Pvu II
Bgl II
30
31
104
105
codon
3'
5'

Figure 1

Figure 2. Comparison of the Nucleotide Sequences of the Rabbit ψβ2 and β1 Globin Genes

The nucleotide sequences of the mRNA synonymous strands of the ψβ2 and β1 (Hardison et al., 1979) genes are aligned. The β1 DNA sequence from the mRNA capping site to the poly(A) addition site, excluding IVS 2, is shown. Insertions/deletions are included in both the ψβ2 and β1 sequences wherever necessary to maintain maximum DNA sequence homology. Vertical lines indicate homologous bases in the two sequences. The ψβ2 sequence is numbered in both directions from the putative cap site. The numbers are placed above the ψβ2 sequence so that the first digit (or the minus sign) is directly above the numbered nucleotide. The numbers beneath the β1 sequence designate amino acid codons of the β1 gene. The boundaries of the intervening sequences (IVS) are designated by arrows. The initiator codon, ATG, the terminator, TGA, and the hexanucleotide, AATAAA, in the 3' untranslated region are underlined in the β1 sequence.

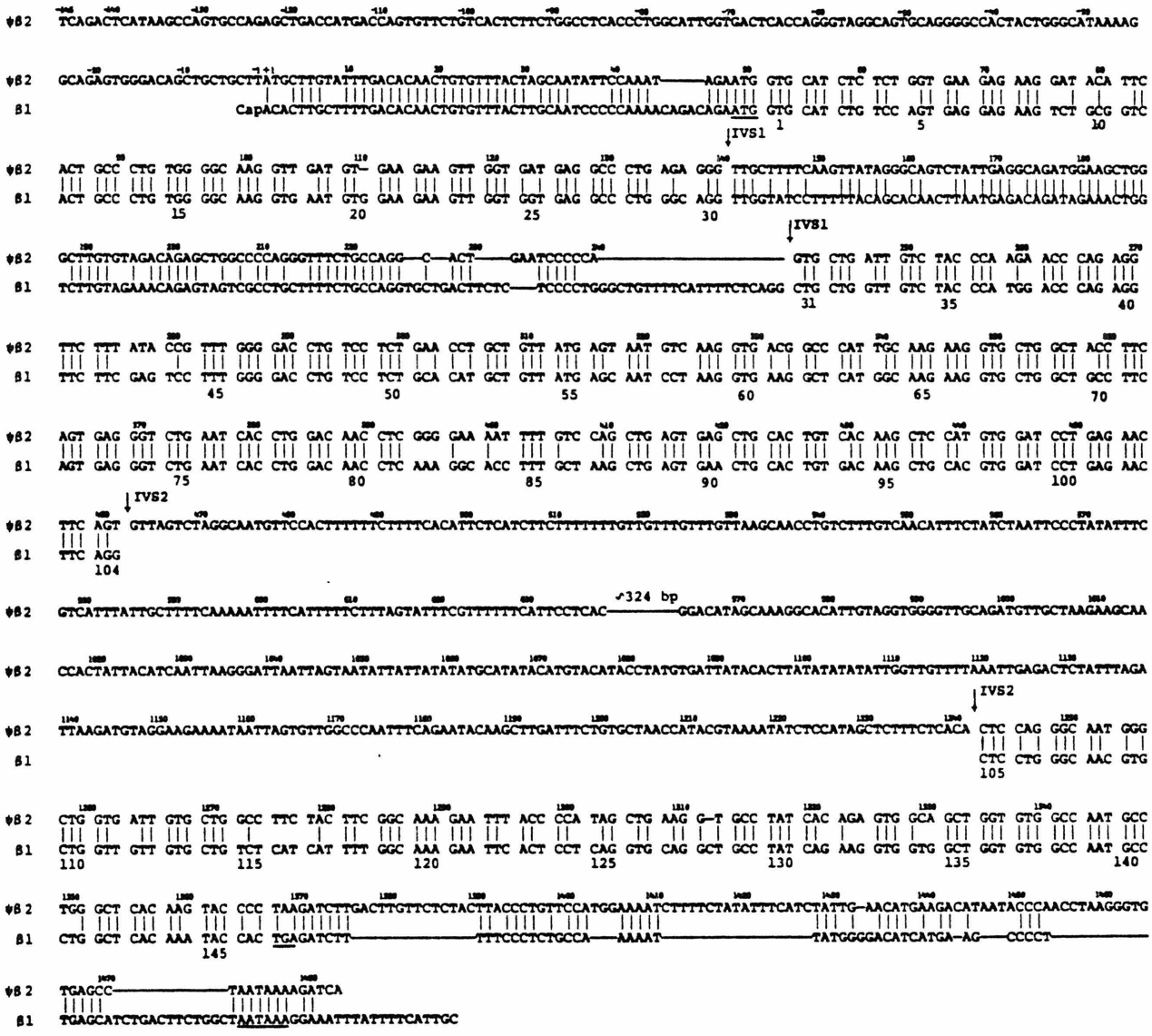

Figure 2

Figure 3. Alignment of the 5' Flanking and Noncoding Sequences of the $\psi\beta 2$, $\beta 1$, Human β and Mouse β^{maj} Globin Genes

In both A and B the arrow denotes the known (or putative $\psi\beta 2$) mRNA capping site. Vertical lines indicate homologous bases in the $\psi\beta 2$ and $\beta 1$ sequences. Vertical lines are also drawn wherever bases are common to the rabbit $\beta 1$, human β and mouse β^{maj} gene sequences. A. An alignment of the $\psi\beta 2$ sequence from base pair -101 to +1 with the sequences 5' to the mRNA capping site in the rabbit $\beta 1$ (Mantei et al., 1979), human β (Lawn et al., 1980), and mouse β^{maj} (Konkel et al., 1978) genes is shown. The ATA and CCAAT sequences are underlined. The lines above the $\psi\beta 2$ sequence denote regions of extensive sequence homology between $\psi\beta 2$ and $\beta 1$. B. An alignment of the $\psi\beta 2$ sequence from base pair 1 to 51 with the sequences from the mRNA capping site through the initiator ATG in the rabbit $\beta 1$, human β and mouse β^{maj} genes is shown. The initiator ATG is underlined.

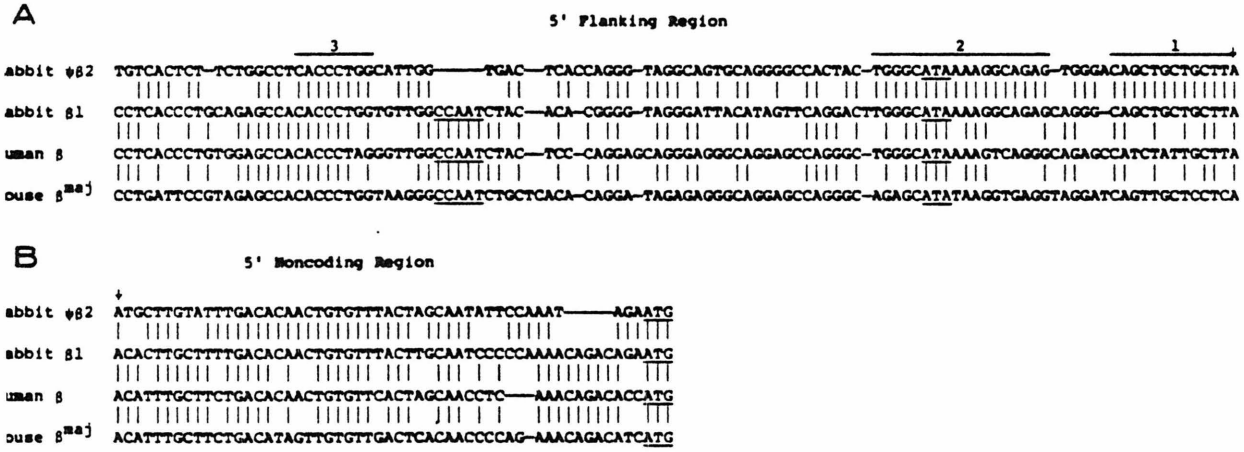
A
5' Flanking Region
3
2
1
abbit ψβ2 TGTCACTCT-TCTGGCCTCACCCTGGCATTGG———TGAC—TCACCAGGG-TAGGCAGTGCAGGGGCCACTAC-TGGGCATAAAAGGCAGAG-TGGGACAGCTGCTGCTTA
abbit β1 CCTCACCCTGCAGAGCCACACCCTGGTGTTGGCCAATCTAC—ACA-CGGGG-TAGGGATTACATAGTTCAGGACTTGGGCATAAAAGGCAGAGCAGGG-CAGCTGCTGCTTA
uman β CCTCACCCTGTGGAGCCACACCCTAGGGTTGGCCAATCTAC—TCC-CAGGAGCAGGGAGGGCAGGAGCCAGGGC-TGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTA
ouse β^maj CCTGATTCCGTAGAGCCACACCCTGGTAAGGGCCAATCTGCTCACA-CAGGA-TAGAGAGGGCAGGAGCCAGGGC-AGAGCATATAAGGTGAGGTAGGATCAGTTGCTCCTCA
B
5' Noncoding Region
abbit ψβ2 ATGCTTGTATTTGACACAACTGTGTTTACTAGCAATATTCCAAAT———AGAATG
abbit β1 ACACTTGCTTTTGACACAACTGTGTTTACTTGCAATCCCCCAAAACAGACAGAATG
uman β ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTC——AAACAGACACCATG
ouse β^maj ACATTTGCTTCTGACATAGTTGTGTTGACTCACAACCCCAG-AAACAGACATCATG

Figure 3

Figure 4. Alignment of Sequences Surrounding Intron/Exon Junctions of ψβ2 and β1

An alignment of the sequences surrounding the four intron/exon junctions of ψβ2 and β1 is shown. Vertical lines denote the splicing sites in β1 which follow the GT/AG rule (Breathnach et al., 1978). Lines are drawn between bases common to the ψβ2 and β1 genes. *The consensus sequences for the 5' and 3' junctions are from Lerner et al. (1980).

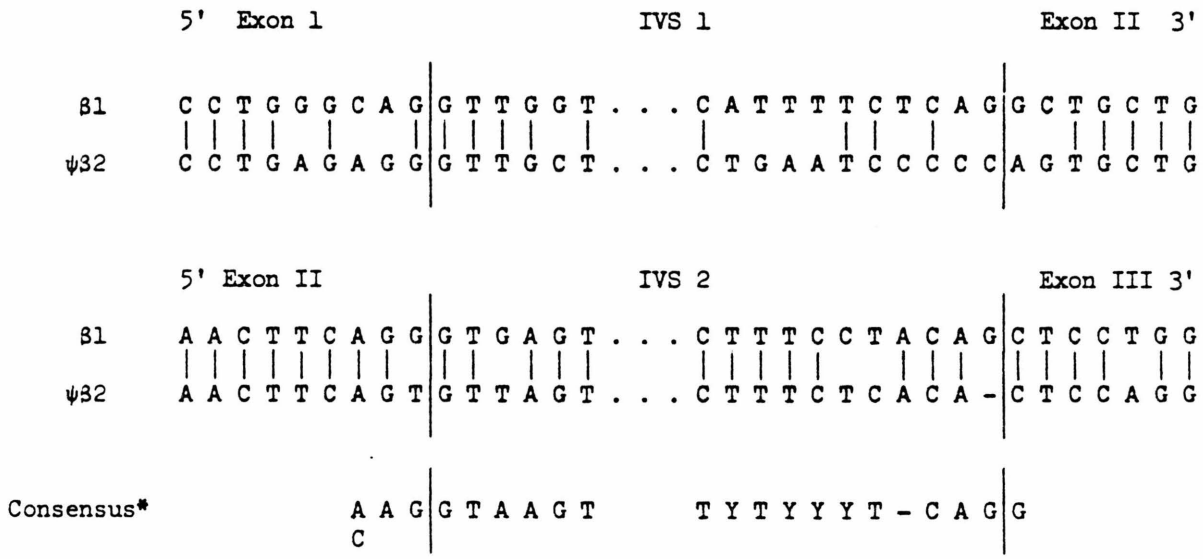
5' Exon 1
IVS 1
Exon II 3'
β1 C C T G G G C A G G T T G G T . . . C A T T T T C T C A G G C T G C T G
ψβ2 C C T G A G A G G G T T G C T . . . C T G A A T C C C C C A G T G C T G
5' Exon II
IVS 2
Exon III 3'
β1 A A C T T C A G G G T G A G T . . . C T T T C C T A C A G C T C C T G G
ψβ2 A A C T T C A G T G T T A G T . . . C T T T C T C A C A - C T C C A G G
Consensus* A A G G T A A G T T Y T Y Y Y T - C A G G
C

Figure 4

Chapter 5

Reprinted from
Proc. Natl. Acad. Sci. USA
Vol. 76, No. 11, pp. 5684–5688, November 1979
Cell Biology

Introduction and expression of a rabbit β-globin gene in mouse fibroblasts

(DNA-mediated gene transfer/cotransformation/intervening sequences/gene regulation)

B. WOLD*, M. WIGLER†, E. LACY‡, T. MANIATIS‡, S. SILVERSTEIN*, AND R. AXEL*

*College of Physicians and Surgeons, Columbia University, New York, New York 10032, †Cold Spring Harbor Laboratories, Cold Spring Harbor, New York 11724, and ‡Division of Biology, California Institute of Technology, Pasadena, California 91125

Communicated by Sol Spiegelman, July 16, 1979

ABSTRACT **The cloned chromosomal rabbit β-globin gene has been introduced into mouse fibroblasts by DNA-mediated gene transfer (transformation). In this report, we examine the expression of the rabbit gene in six independent transformants that contain from 1 to 20 copies of the cloned globin gene. Rabbit globin transcripts were detected in two of these transformants at steady-state concentrations of 5 and 2 copies per cell. The globin transcripts from one cell line are polyadenylylated and migrate as 9S RNA on methylmercury gels. These transcripts reflect correct processing of the two intervening sequences but lack 48 ± 5 nucleotides present at the 5′ terminus of rabbit erythrocyte globin mRNA.**

Cellular genes coding for selectable biochemical functions can be stably introduced into cultured mammalian cells by DNA-mediated gene transfer (transformation) (1, 2). Biochemical transformants are readily identified by the stable expression of a gene coding for a selectable marker. These transformants represent a subpopulation of competent cells that integrate other physically unlinked genes for which no selective criteria exist (3). In this manner, we have used a viral thymidine kinase (tk) gene as a selectable marker to isolate mouse cell lines that are stably transformed with the tk gene along with bacteriophage φX174, plasmid pBR322, or the cloned chromosomal rabbit β-globin gene sequences (3).

Cotransformed mouse fibroblasts containing the rabbit β-globin gene provide an opportunity to study the expression and subsequent processing of these sequences in a heterologous host. In this report, we demonstrate the expression of the transformed rabbit β-globin gene generating a discrete polyadenylylated 9S species of globin RNA. This RNA results from correct processing of both intervening sequences, but lacks approximately 48 nucleotides present at the 5′ terminus of mature rabbit β-globin mRNA.

MATERIALS AND METHODS

Cell Culture. Murine Ltk⁻ aprt⁻ cells are adenine phosphoribosyltransferase-negative derivatives of Ltk⁻ clone 1D cells (4) that were originally isolated and characterized by R. Hughes and P. Plagemann. Cells were maintained in growth medium and prepared for transformation as described (5).

Transformation and Selection. The transformation protocol, selection for tk⁺ transformants, and maintenance of transformant cell lines were as described (5).

DNA Isolation. DNA was extracted from cultured L cells as described (5). Recombinant phage containing the rabbit β-globin gene in the λ phage vector Charon 4A were grown and purified, and DNA was isolated as described (6). The herpes virus DNA fragment containing the tk gene was purified from total DNA of herpes simplex virus strain F (7). Intact herpes virus DNA was digested with the restriction endonuclease *Kpn* I and fractionated by agarose gel electrophoresis, and the 5.1-kilobase pair (kbp) fragment containing the tk gene was extracted from the gel as described (8).

RNA Isolation. Total RNA was isolated from logarithmic-phase cultures of transformed L cells by successive extractions with phenol at pH 5.1, phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol), and chloroform/isoamyl alcohol (24:1, vol/vol). After ethanol precipitation, the RNA was digested with DNase (9) and precipitated with ethanol. Nuclear and cytoplasmic fractions were isolated as described (5) and RNAs were extracted as described above. Cytoplasmic polyadenylylated RNA was isolated by oligo(dT)-cellulose chromatography (10).

cDNA Synthesis. Rabbit and mouse cDNAs were prepared by using avian myeloblastosis virus reverse transcriptase (RNA-dependent DNA polymerase) (obtained from J. W. Beard), as described (11).

DNA Filter Hybridizations. Cellular DNA was digested with restriction endonucleases, electrophoresed on agarose slab gels, transferred to nitrocellulose filter sheets, and hybridized with ^{32}P-labeled DNA probes as described by Wigler *et al.* (5).

Solution Hybridizations. ^{32}P-Labeled globin cDNAs (specific activities of 2.9×10^8 cpm/μg) were hybridized with excess RNA in 0.4 M NaCl/25 mM 1,4-piperazinediethanesulfonic acid (Pipes), pH 6.5/5 mM EDTA at 75° C. Incubation times did not exceed 70 hr. R_0ts were calculated as moles of RNA nucleotides per liter times time in seconds. The fraction of cDNA rendered resistant to the single-strand nuclease S1 in hybridization was determined as described (10).

RNA Filter Hybridizations. RNA was electrophoresed through 1% agarose slab gels (17 × 20 × 0.4 cm) containing 5 mM methylmercury hydroxide as described by Bailey and Davidson (12). The concentration of RNA in each slot was 0.5 μg/μl. Electrophoresis was at 110 V for 12 hr at room temperature.

RNA was transferred from the gel to diazotized cellulose paper as described by Alwine *et al.* (13) by using pH 4.0 citrate transfer buffer. After transfer, the RNA filter was incubated for 1 hr with transfer buffer containing carrier RNA at 500 μg/ml. The RNA on the filters was hybridized with cloned DNA probe at 50 ng/ml labeled by ^{32}P nick translation (14) to specific activities of $2–8 \times 10^8$ cpm/μg. Reaction volumes were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: tk, thymidine kinase; kbp, kilobase pairs; Pipes, 1,4-piperazinediethanesulfonic acid; R_0t, product of RNA concentration (moles of nucleotide per liter) and incubation time (seconds).

5 μl/cm² of filter. Hybridization was in 4× standard saline trate (0.15 M NaCl/0.015 M sodium citrate)/50% formamide 57°C for 36–48 hr.

After hybridization, filters were soaked in two changes of 2× andard saline citrate/25 mM sodium phosphate/1.5 mM so- um pyrophosphate/0.1% sodium dodecyl sulfate/5 mM DTA at 37°C for 30 min with shaking to remove formamide. ıccessive washes were at 68°C with 1× and 0.1× standard line citrate containing 5 mM EDTA and 0.1% sodium dodecyl lfate for 30 min each.

Berk–Sharp Analysis of Rabbit β-Globin RNA in Trans- rmed Mouse L Cells. The hybridizations were carried out 80% (vol/vol) formamide (Eastman)/0.4 M Pipes, pH 6.5/0.1 M EDTA/0.4 M NaCl (15, 16) for 18 hr at 51°C for the 1.8)p *Hha* I fragment and 49°C for the *Pst* I fragment. The /brids were treated with S1 nuclease and analyzed by a odification of the procedure described by Berk and Sharp 6).

RESULTS

ansformation of mouse cells with the rabbit globin gene

e have performed cotransformation experiments with the romosomal adult rabbit β-globin gene, using the purified rpes virus tk gene as a biochemical marker. The addition of e tk gene to mutant Ltk⁻ mouse fibroblasts results in the ap- arance of stable transformants that can be selected by their ility to grow in hypoxanthine/aminopterin/thymidine AT) medium. Cells were cotransformed with a β-globin gene ne designated RβG1, which consists of a 15.5-kbp insert of bbit DNA carried in the bacteriophage λ cloning vector Charon 4A (Fig. 1*A*) (unpublished data). The purified tk gene was mixed with a 100-fold molar excess of intact recombinant DNA from clone RβG1. This DNA was then exposed to mouse Ltk⁻ cells under transformation conditions previously described (5). After 2 weeks in selective medium, tk⁺ transformants were observed at a frequency of one colony per 10^6 cells per 20 pg of tk gene. Clones were picked and grown into mass culture.

We then asked if the tk⁺ transformants also contain rabbit β-globin sequences. High molecular weight DNA from eight transformants was cleaved with the restriction endonuclease *Kpn* I. The DNA was fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters, and these filters were then annealed with nick-translated globin [^{32}P]DNA [blot hybridization (17)]. Cleavage of this recombinant phage with the enzyme *Kpn* I generates a 4.7-kbp fragment that contains the entire adult β-globin gene, along with 1.4 kbp of 5′ flanking information and 2.0 kbp of 3′ flanking information (Fig. 1*A*). This fragment was purified by gel electrophoresis and nick translated to generate a hybridization probe. Blot hybridization experiments showed that the 4.7-kbp *Kpn* I fragment containing the globin gene was present in the DNA of six of the eight tk⁺ transformants. In three of the clones (Fig. 2, lanes E, F, and H), additional rabbit globin bands were observed, which probably resulted from the loss of at least one of the *Kpn* I sites during transformation. The number of rabbit globin genes integrated in these transformants was variable: some clones contained a single copy of the gene (Fig. 2, lanes J and K), whereas others contained up to 20 copies of the heterologous gene. It should be noted that the β-globin genes of mouse and rabbit are partially homologous. However, we do not observe hybridization of the rabbit β-globin probe to *Kpn*-cleaved mouse DNA, presumably because *Kpn* cleavage of mouse DNA leaves the β-gene cluster in exceedingly high molecular weight fragments not readily detected in these experiments (Fig. 2). These results demonstrate the introduction of the cloned chromosomal rabbit β-globin gene into mouse cells by DNA-mediated gene transfer.

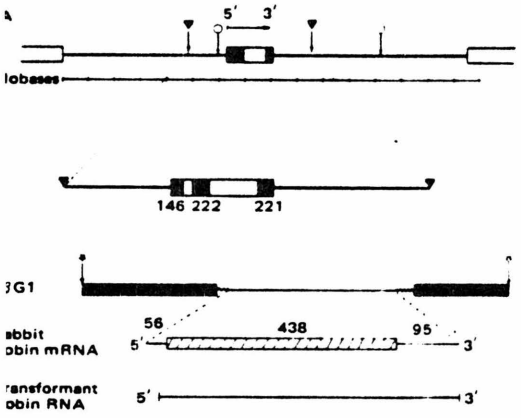

FIG. 1. (*A*) Structure of the rabbit β-globin genomic clone RβG1. ıe solid box represents the mRNA coding sequence in the adult globin gene. The clear regions bounded by coding sequence indicate intervening sequences within the β-globin gene. The larger 3′ ervening sequence is about 600 base pairs long and the smaller 5′ uence (shown only in the lower map) is about 125 base pairs long. striction sites are indicated by arrows: ▼, *Kpn* I; ○, *Pst* I. (*B*) ructure of the cDNA clone pβG1 and rabbit β-globin mRNA. The a I restriction fragment of pβG1 is shown. The heavy black lines licate pMB9 plasmid vector sequence and the thin straight line licates rabbit mRNA sequence: ☆ *Hha* I sites. The map of rabbit bin mRNA shows the 438-nucleotide translated region bounded the 5′ 56-nucleotide untranslated region and the 3′ 95-nucleotide translated region. The bottom map is of cytoplasmic polyadenylted rabbit globin RNA from transformant cell line 6, which lacks proximately 48 nucleotides of 5′ mRNA sequence (see *Results*).

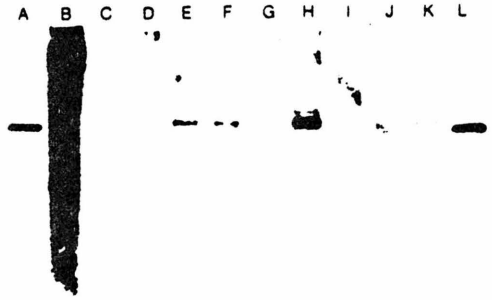

FIG. 2. Rabbit β-globin genes in transformed mouse L cells. High molecular weight DNA from eight independent cotransformant clones was digested with *Kpn* I and electrophoresed on a 0.7% agarose gel. The DNA was denatured *in situ* and transferred to nitrocellulose filters, which were then annealed with a [^{32}P]-labeled 4.7-kbp fragment containing the rabbit β-globin gene. Lanes A and L, 50 pg of the 4.7-kbp *Kpn* fragment of RβG1; lane B, 15 μg of rabbit liver DNA digested with *Kpn*; lane C, 15 μg of Ltk⁻ aprt⁻ DNA; lanes D–K, 15 μg of DNA from each of eight independently isolated tk⁺ transformants.

ıbbit β-globin sequences are transcribed in mouse ınsformants

ıe cotransformation system we have developed may provide unctional assay for cloned eukaryotic genes if these genes are pressed in the heterologous recipient cell. Six transformed ll clones were therefore analyzed for the presence of rabbit globin RNA sequences. In initial experiments we performed ution hybridization reactions to determine the cellular ncentration of rabbit globin transcripts in our transformants. radioactive cDNA copy of purified rabbit α- and β-globin RNA was annealed with a vast excess of cellular RNA. Beuse homology exists between the mouse and rabbit globin quences, it was necessary to determine experimental condiıns such that the rabbit globin cDNAs did not form stable 'brids with mouse globin mRNA but did react completely ıth homologous rabbit sequences. At 75°C in the presence of 4 M NaCl, over 80% hybridization was observed with the bbit globin mRNA, whereas the heterologous reaction with ırified mouse globin mRNA did not exceed 10% hybridizaın. The $R_0t_{1/2}$ of the homologous hybridization reaction was $\times 10^{-4}$, a value consistent with a complexity of 1250 nucleides contributed by the α- plus β-globin sequences in our)NA probe (10).

This rabbit globin cDNA was used as a probe in hybridization actions with total RNA isolated from six transformed cell lines 'ig. 3 and data not shown). Total RNA from transformed clone (Fig. 2, lane H) protected 44% of the rabbit cDNA at cometion, the value expected if only β-gene transcripts were 'esent. This reaction displayed pseudo-first-order kinetics with ı $R_0t_{1/2}$ of 2×10^3. A second transformant (Fig. 2, lane E) acted with an $R_0t_{1/2}$ of 8×10^3 (data not shown). No signifiınt hybridization was observed at R_0ts $\geq 10^4$ with total RNA reparations from the four additional transformants.

We have characterized the RNA from clone 6 in greatest etail. RNA from this transformant was fractionated into nuear and cytoplasmic populations to determine the intracellular calization of the rabbit globin RNA. The cytoplasmic RNA

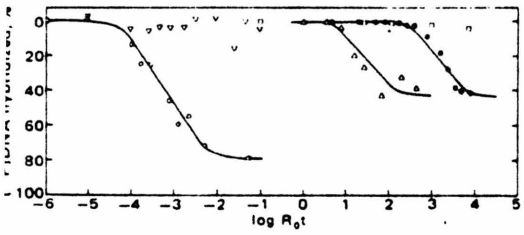

FIG. 3. Hybridization of rabbit globin cDNA with RNA from a ıouse L cell transformant containing RβG1 sequences. The curves ›present single pseudo-first-order kinetic components fit to the data y least-squares methods. ○, Hybridization of rabbit globin ^{32}P]cDNA (7–9 × 10^8 cpm/μg) with excess globin template RNA. At ermination, 80% of the cDNA is in hybrid. The rate constant is 1.1 $\times 10^3$ M^{-1} sec^{-1}. ▽, Hybridization of rabbit globin cDNA with mouse lobin mRNA. △, Hybridization of excess polyadenylylated cytolasmic RNA from transformant 6 (see text) with rabbit globin cDNA. 'he rate constant is 2.8×10^{-2} M^{-1} sec^{-1}. The extent of reaction was 3% after normalization for the 70% reactivity of the cDNA at the time f this measurement. ●, Hybridization of excess total cellular RNA rom transformant 6 with rabbit globin cDNA. At termination, 43% f the [^{32}P]cDNA was in hybrid. The rate constant is 13.5×10^{-4} M^{-1} ec^{-1}. □, Hybridization of excess nuclear RNA from transformant with rabbit globin cDNA. The S1 resistance of cDNA at zero time ıas been subtracted from all hybridization values. These background 'alues were 5% and 14% for the cDNA preparations used in this exeriment.

was further fractionated by oligo(dT)-cellulose chromatography into poly(A)$^+$ and poly(A)$^-$ RNA. Poly(A)$^+$ cytoplasmic RNA from clone 6 hybridizes with the rabbit cDNA with an $R_0t_{1/2}$ of 25. This value is 1/80th the $R_0t_{1/2}$ observed with total cellular RNA, consistent with the observation that poly(A)$^+$ cytoplasmic RNA is 1–2% of the total RNA in a mouse cell. Hybridization is not detectable with either nuclear RNA or cytoplasmic poly(A)$^-$ RNA at R_0t values of 1×10^4 and 2×10^4, respectively. The steady-state concentration of rabbit β-globin RNA present in our transformant can be calculated from the $R_0t_{1/2}$ to be about five copies per cell, with greater than 90% localized in the cytoplasm.

Several independent experiments argue that the globin RNA detected derives from transcription of the rabbit DNA sequences present in this transformant: (*i*) cDNA was prepared from purified 9S mouse globin RNA. This cDNA does not hybridize with poly(A)$^+$ RNA from clone 6 at R_0t values at which the reaction with rabbit globin cDNA is complete (Fig. 3). (*ii*) Rabbit globin cDNA does not hybridize with total cellular RNA obtained with tk$^+$ globin$^-$ transformants at R_0t values exceeding 10^4. (*iii*) The hybridization we observe does not result from duplex formation with rabbit globin DNA possibly contaminating the RNA preparations. Rabbit cDNA was annealed with total cellular RNA from clone 6, the reaction product was treated with S1 nuclease, and the duplex was subjected to equilibrium density centrifugation in cesium sulfate under conditions that separate DNA·RNA hybrids from duplex DNA. The S1-resistant cDNA banded at a density of 1.54 g/ml, as expected for DNA·RNA hybrid structures (data not shown). These data, along with the observation that globin RNA is polyadenylylated, demonstrate that the hybridization we observe with RNA preparations does not result from contaminating DNA sequences.

Characterization of rabbit globin transcripts in transformed cells

In rabbit erythroblast nuclei, the β-globin gene sequences are detected as a 14S precursor RNA that reflects transcription of two intervening sequences that are subsequently removed from this molecule to generate a 9S messenger RNA (unpublished results). It was therefore of interest to determine whether the globin transcripts we detected exist as a discrete 9S species, which is likely to reflect appropriate splicing of the rabbit gene transcript by the mouse fibroblast. Cytoplasmic poly(A)-containing RNA from clone 6 was electrophoresed on a methylmercury/agarose gel (12) and transferred to diazotized cellulose paper (13, 18). After transfer, the RNA on the filters was hybridized with DNA from the plasmid pβG1, which contains rabbit β-globin cDNA sequences (19). Using this ^{32}P-labeled probe, we observed a discrete 9S species of RNA in the cytoplasm of the transformant, which comigrated with rabbit globin mRNA isolated from rabbit erythroblasts (Fig. 4). Hybridization to 9S RNA species was not observed in parallel lanes containing either purified mouse 9S globin RNA or poly(A)-containing cytoplasmic RNA from a tk$^+$ transformant containing no rabbit globin genes.

We were unable in these experiments to detect the presence of a 14S precursor in nuclear RNA populations from the transformants. This is not surprising, because the levels expected in nuclear RNA, given the observed cytoplasmic concentration, are likely to be below the limits of detection for this technique. The 5′ and 3′ boundaries of the rabbit globin sequences expressed in transformed fibroblasts along with the internal processing sites can be defined more accurately by hybridizing this RNA with cloned DNAs, followed by S1 nuclease digestion and subsequent gel analysis of the DNA products (16). When

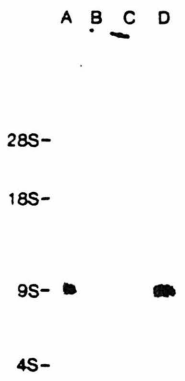

FIG. 4. Sizing of cytoplasmic polyadenylylated rabbit globin ıanscripts from transformant 6. RNA was electrophoresed in a 1% ethylmercury/agarose gel and the RNA was transferred to diazo-ed cellulose paper. The positions of 28S, 18S, and 4S RNAs on the l were determined optically after staining with ethidium bromide. ıe RNA on the filter was hybridized with ^{32}P-labeled plasmid DNA pβG1) containing the rabbit β-globin cDNA sequence. Lane A, 1 ng purified 9S polyadenylylated RNA from rabbit reticulocytes, plus μg of carrier chicken oviduct RNA. Lane B, 50 pg of purified 9S dyadenylylated RNA from rabbit reticulocytes, plus 25 μg of carrier icken oviduct RNA. Lane C, 1 ng of purified 9S polyadenylylated NA from mouse reticulocytes plus 25 μg of carrier RNA. Lane D, μg of polyadenylylated cytoplasmic RNA from transformant 6. ne E, 30 μg of cytoplasmic polyadenylylated RNA from a trans-rmant containing no rabbit globin genes.

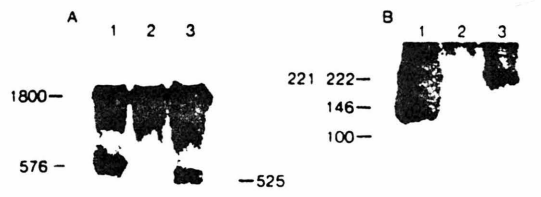

FIG. 5. Characterization of rabbit β-globin RNA in transformed mouse L cells. Numbers of base pairs are given beside the autora diograms. (*A*) Both total rabbit reticulocyte RNA and poly(A)$^+$ RNA purified from cell line 6 were hybridized to the 1.8-kbp *Hha* I fragment from plasmid pβG1 (Fig. 1*B*) and analyzed as described by Berk and Sharp (16). Lane 1, 0.2 μg of total reticulocyte RNA was hybridized to 20 ng of the 1.8-kbp *Hha* globin fragment in 5 μl. Lane 2, 18 ng of the 1.8-kbp *Hha* fragment was hybridized in 2.5 μl in the absence of any added RNA. Lane 3, 30 μg of poly(A)$^+$ RNA purified from cell line 6 was hybridized to 75 ng of the 1.8 kbp *Hha* fragment in 10 μl. The 1800-base pair band is the renatured *Hha* fragment. (*B*) Both total rabbit reticulocyte RNA and poly(A)$^+$ RNA purified from cell line 6 were hybridized to a 5.60-kbp *Pst* I fragment containing the genomic copy of the rabbit β-globin gene. The Berk–Sharp analysis was carried out by a procedure to be described elsewhere. Only the bottom half of the autoradiogram is shown and therefore lane-specific background present in lanes 1 and 3, as well as in the RNA control (lane 2) is not shown. We believe that this background results from the formation of DNA·DNA duplexes between a small number of nicked *Pst* fragments prior to S1 treatment. Lane 1, 0.35 μg of total rabbit reticulocyte RNA was hybridized to 0.1 μg of the 5.60 kbp *Pst* fragment in 10 μl. Lane 2, 0.12 μg of the *Pst* fragment was hybridized in 10 μl in the absence of any RNA. Lane 3, 30 μg of poly(A)$^+$ RNA purified from cell line 6 was hybridized to 0.12 μg of the 5.60 kbp *Pst* fragment in 10 μl.

-globin mRNA from rabbit erythroid cells was hybridized ith cDNA clone pβG1 (Fig. 1*B*) under appropriate conditions, ıe entire 576-base pair insert of cDNA was protected from S1 ıclease attack. When this cDNA clone was hybridized with NA from our transformant, surprisingly, a discrete DNA band as observed at 525 base pairs, but not at 576 base pairs (Fig.). These results suggest that, in this transformant, rabbit globin NA molecules are present that have a deletion in a portion of ıe globin mRNA sequence at the 5′ or 3′ termini. To distin-uish between these possibilities, DNA of the λ clone, Rβ*G*1, ontaining the chromosomal rabbit β-globin sequence hy-ridized with transformed fibroblast RNA. The hybrid formed as treated with S1 nuclease, and the protected DNA fragments ere analyzed by alkaline agarose gel electrophoresis and lentified by Southern blotting procedures (17). Because the ıbbit β-globin gene is interrupted by two intervening se-uences, the hybridization of mature rabbit mRNA to Rβ*G*1 NA generates three DNA fragments in this sort of analysis: 146-base pair fragment spanning the 5′ terminus to the ınction of the small intervening sequence, a 222-base pair ıternal fragment bridging the small and large intervening equences, and a 221-base pair fragment spanning the 3′ ınction of the large intervening sequence to the 3′ terminus f the mRNA molecule (Fig. 1*A*). When transformant RNA was nalyzed in this fashion, we observed a 222-base pair fragment nd an aberrant fragment of 100 base pairs but no 146-base pair ragment (Fig. 5). Hybridization with a specific 5′ probe ıowed that the internal 222 base pair fragment was present data not shown). The sum of the protected lengths equaled the ngth of the DNA fragment protected by using the cDNA lone. Taken together, these results indicate that although the ıtervening sequences expressed in transformed mouse fibro-last are removed from the RNA transcripts precisely, the 5′ ermini of the cytoplasmic transcripts we observe do not contain about 48 ± 5 nucleotides present in mature 9S RNA of rabbit erythroblasts.

DISCUSSION

In these studies, we have constructed mouse cell lines that contain the rabbit β-globin gene and have analyzed the ability of the mouse fibroblast recipient to transcribe and process this heterologous gene. Solution hybridization experiments in concert with RNA blotting techniques indicate that, in at least one transformed cell line, rabbit globin sequences are expressed in the cytoplasm as a polyadenylylated 9S species. Correct processing of the rabbit β-globin gene has also been observed in tk$^+$ mouse cell transformants in which the globin and tk plasmids have been ligated prior to transformation (20). Similar results have been obtained by using a viral vector to introduce the rabbit globin gene into monkey cells (21, 22). Taken together, these results suggest that nonerythroid cells from heterologous species contain the enzymes necessary to correctly process the intervening sequences of a rabbit gene whose expression usually is restricted to erythroid cells.

The level of expression of rabbit globin sequences in our transformant is low: five copies of globin RNA are present in the cytoplasm of each cell. Our results indicate that the two intervening sequences present in the original globin transcript are processed and removed at loci indistinguishable from those observed in rabbit erythroid cells. Surprisingly, 45 nucleotides present at the 5′ terminus of mature rabbit mRNA are absent

from the β-globin RNA sequence detected in the cytoplasm of the transformant we have examined. It is possible that incorrect initiation of transcription occurs about the globin gene in this mouse cell line. Alternatively, the globin sequences we detect may result from transcription of a long precursor that ultimately must undergo 5′ processing to generate the mature 9S species. Incorrect processing at the 5′ terminus in the mouse fibroblast could be responsible for our results. At present, it is difficult to distinguish among these alternatives. Because we are restricted in our analysis to a single transformant, we do not know whether these observations are common to all transformants expressing the globin gene or reflect a rare but interesting abberation. It should be noted, however, that in similar experiments by Weissmann and his colleagues (20) at least a portion of the rabbit globin RNA molecules transcribed in transformed mouse fibroblasts retain the correct 5′ terminus.

Several alternative explanations can be offered for the expression of globin sequences in transformed fibroblasts. It is possible that constitutive synthesis of globin RNA occurs in cultured fibroblasts (23) at levels five to six orders of magnitude below the level observed in erythroblasts. The introduction of 20 additional globin DNA templates may simply increase this constitutive transcription to the levels observed in our transformant. Alternatively, it is possible that the homologous globin gene is repressed by factors that are partially overcome by a gene dosage effect provided by the introduction of 20 additional globin genes. Finally, normal repression of the globin gene in a fibroblast may depend upon the position of these sequences in the chromosome. At least some of the newly introduced genes are likely to reside at loci distant from the resident mouse globin genes. Some of these ectopic sites may support low level transcription. Our data do not permit us to distinguish among these and other alternatives.

Although the number of rabbit globin genes within a given transformant remains stable for over a hundred generations of culture in hypoxanthine/aminopterin/thymidine (unpublished studies), it has not been possible to prove that these sequences are covalently integrated into recipient cell DNA. In previous studies, however, we have demonstrated that cotransformation of either φX174 or plasmid pBR322 results in the stable integration of these sequences into high molecular nuclear DNA. In the present study, the globin gene represents a small internal segment of the high molecular weight concatenated phage DNA used in the transformation (Fig. 1A). Analysis of integration sites covalently linked to donor DNA is therefore difficult. Preliminary studies using radioactive λ sequences as a probe in DNA blotting experiments indicate that, in some of our cell lines, we have introduced a contiguous stretch of recombinant phage DNA with a minimum length of 50 kbp.

The presence of 9S globin RNA in the cytoplasm of transformants suggests that this RNA may be translated to give rabbit β-globin polypeptide. Attempts to detect this protein in cell lysates using a purified anti-rabbit β-globin antibody (kindly provided by S. Boyer) have thus far been unsuccessful. It is possible that the globin RNAs in our transformant are not translated or are translated with very low efficiency due to the absence of a functional ribosomal binding site. The cytoplasmic globin transcripts in our transformant lack about 48 nucleotides of untranslated 5′ sequence (Fig. 1B), which includes 13 nucleotides known to interact with the 40S ribosomal subunit in nuclease protection studies (24, 25). Even if translation did occur with normal efficiency, it is probable that the protein would exist at levels below the limits of detection of our immunologic assay due to the low level of globin RNA, and the observation that the half-life of β globin in the absence of heme and α globin may be less than 30 min (22).

These studies indicate the potential value of cotransformation systems in the analysis of eukaryotic gene expression. The introduction of wild-type genes along with native and *in vitro*-constructed mutant genes into cultured cells provides an assay for the functional significance of sequence organization. It is obvious from these studies that this analysis will be facilitated by the ability to extend the generality of cotransformation to recipient cell lines, such as murine erythroleukemia cells, that provide a more appropriate environment for the study of heterologous globin gene expression.

1. Wigler, M., Silverstein, S., Lee, L. S., Pellicer, A., Cheng, Y. C. & Axel, R. (1977) *Cell* **11**, 223–232.
2. Maitland, N. J. & McDougall, J. K. (1977) *Cell* **11**, 233–241.
3. Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. & Axel, R. (1979) *Cell* **16**, 777–785.
4. Kit, S., Dubbs, D., Piekarski, L. & Hsu, T. (1963) *Exp. Cell Res.* **31**, 291–312.
5. Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. & Chasin, L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1373–1376.
6. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
7. Pellicer, A., Wigler, M., Axel, R. & Silverstein, S. (1978) *Cell* **14**, 133–141.
8. Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. (1978) *Cell* **14**, 725–731.
9. Maxwell, I. H., Maxwell, F. & Hahn, W. E. (1977) *Nucleic Acids Res.* **4**, 241–246.
10. Axel, R., Feigelson, P. & Schutz, G. (1976) *Cell* **7**, 247–254.
11. Myers, J. C. & Spiegelman, S. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5329–5333.
12. Bailey, J. & Davidson, N. (1976) *Anal. Biochem.* **70**, 75–85.
13. Alwine, J., Kemp, D., Parker, B., Reiser, J., Renart, J., Stark, G. & Wahl, G. (1979) *Methods Enzymol.* **68**, in press.
14. Wellstock, R., Sweet, R., Weiss, M., Cedar, H. & Axel, R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1299–1303.
15. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4**, 1539–1552.
16. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
17. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
18. Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5350–5354.
19. Maniatis, T., Sim, G. K., Efstratiadis, A. & Kafatos, F. C. (1976) *Cell* **8**, 163–182.
20. Mantei, N., Boll, W. & Weissmann, C. (1979) *Nature (London)* **281**, 40–46.
21. Hamer, D. H. & Leder, P. (1979) *Nature (London)*, **281**, 35–39.
22. Mulligan, R. C., Howard, B. H. & Berg, P. (1979) *Nature (London)* **277**, 108–114.
23. Humphries, S., Windass, J. & Williamson, R. (1976) *Cell* **7**, 267–277.
24. Efstratiadis, A., Kafatos, F. & Maniatis, T. (1977) *Cell* **10**, 571–585.
25. Legon, S. (1976) *J. Mol. Biol.* **106**, 37–53.