- R. Wells, H. D. Royer, C. P. Hollenberg, *Mol. Gen. Genet.* **147**, 45 (1976).
 E. H. Davidson and R. J. Britten, *Q. Rev. Biol.* **48**, 565 (1973); E. H. Davidson, W. H. Klein, R. H. Davidson, W. H. Klein, R. H. Davidson, W. H. Klein, R. H. Davidson, M. H. Klein, R. H. Barten, Go (1977).
- J. Britten, Dev. Biol. 55, 69 (1977).
 M. So, R. Gill, S. Falkow, Mol. Gen. Genet. 142, 239 (1975). Our work was done in compliance with the NIH guidelines for recombinant DNA waves of the second secon research.
- research.
 R. Roychoudhurg, E. Jay, R. Wu, *Nucleic Acids Res.* 3, 101 (1976).
 R. W. Davis, M. Simon, N. Davidson, *Methods Enzymol.* 21D, 413 (1971).
 D. M. Glover, R. L. White, D. J. Finnegan, D. S. Hogness, *Cell* 5, 149 (1975).
 G. S. Hayward, *Virology* 49, 342 (1972).

- R. J. Britten, D. E. Graham, F. C. Eden, D. M. Painchaud, E. H. Davidson, J. Mol. Evol. 9, 1 (1976).
- 16. A. S. Lee and R. L. Sinsheimer, Proc. Natl. Acad. Sci. U.S.A. 71, 2882 (1974). W. Barnes, unpublished observations.
- 18. Five of these clones were previously studied by W. R. Crain in this laboratory. The results of both
- Subjects are in close agreement. Supported by NIH grants HD 05753 and GM 20927, and by NSF grant BMS 75-07359. A.S.L. has been supported by a fellowship from the American Cancer Society, California Division, J-289. studies are in close agreement 19.

7 February 1977

Nucleotide Sequences from a Rabbit Alpha Globin Gene **Inserted in a Chimeric Plasmid**

Abstract. Rabbit alpha globin gene copies have been made, using reverse transcriptase and DNA polymerase I, and cloned in bacterial plasmids. Plasmid pHb72 has been shown to contain the alpha gene sequence by restriction enzyme analysis and nucleotide sequencing studies, and therefore has been approved for propagation under P2 plus EK1 conditions by the National Recombinant DNA Committee.

Reverse transcriptase has been used to synthesize complementary DNA (cDNA) copies of a variety of polyadenylated messenger RNA's (mRNA's). Several years ago, we discovered that reverse transcriptase leaves a short doublestranded hairpin structure when it terminates the synthesis of such cDNA's. This structure provides an efficient primer for second-strand synthesis by DNA polymerase I(l) so that it is possible to synthesize a double-stranded DNA copy of the mRNA sequence. This, and similar observations (2), prompted us as well as others to use the self-priming characteristics of cDNA's made with reverse transcriptase to develop general methods for cloning sequences copied from mammalian mRNA's in bacterial plasmids (3-5). Other methods have been developed for cloning mRNA sequences that did not utilize the self-priming nature of cDNA's synthesized by reverse transcriptase (6). In some of these cases, however, the use of complicated steps to force priming artificially at the desired location resulted in cloning efficiencies one to two orders of magnitude lower than that which we reported.

We have described the characterization of rabbit beta globin clones that we constructed using the cDNA technique (3, 7). Our analysis included the determination of a 169-nucleotide sequence coding for the beta globin amino acid sequences from cysteine 93 through the carboxyl terminus of the structural gene (7, 8). In the course of screening a series of 36 plasmids (pHb51 through pHb86) for the presence and the orientation of beta globin gene sequences, we identified nine plasmids that appeared to contain alpha globin gene sequences. Here we report the preliminary characterization of these plasmids.

Digestion with the two restriction endonucleases Eco RI and Bam HI provides a convenient method for screening beta globin gene sequences. A fragment 70 nucleotides in length is liberated from plasmids which carry that portion of the beta globin gene which codes for amino acid residues 98 through 122. In addition, it is possible to determine rapidly the orientation of beta globin gene sequences within these plasmids by comparing the fragments liberated by digestion with Bam HI with those liberated with both Bam HI and Eco RI. Since there is a Bam HI site within the beta globin gene and another within the plasmid gene for tetracycline resistance. Bam HI digestion liberates a fragment in the size range 600 to 1100 base pairs. If the orientation of the beta gene insertion is such that the Eco RI site is between the two Bam HI sites, then the corresponding fragment from a Bam HI plus Eco RI digestion will be 70 base pairs shorter. With the reverse orientation the Eco RI digestion does not affect this fragment.



Fig. 1. Plasmid pHb72 restriction fragment map. The number pairs above each line indicate amino acid correspondence for sequences bearing confirmed sites for the restriction endonucleases cited. The number pairs under each line indicate amino acid correspondence for sequences bearing possible sites for the enzymes cited. Single numbers between sites indicate the number of nucleotides between sites. The inserted sequences in pHb72 extend from the codon for valine (Val) 121 to approximately 12 nucleotides beyond the initiation codon. A wavy line indicates poly(dA dT) homopolymer. The lengths shown (numbers in parentheses) for these homopolymer regions are approximate and were estimated from ladder sequencing gels (data not shown). The arrangement of the fragments in the map was determined by a combination of conventional techniques and correlation of sequencing results with the known amino acid sequence (10).

Of the 13 plasmids assayed in this manner, nine of the inserts were oriented with the 3' end (mRNA sense) nearest the external pMB9 Bam HI site. Four were oriented in the opposite direction. This is an interesting contrast to the results of Maniatis *et al.* (4), who found that of 19 beta globin clones examined, all were oriented with the 3' end nearest the external Bam HI site.

In the course of such studies it was observed that plasmid pHb72 contained an Eco RI site but no Bam HI site within the plasmid insert. Such a result might be expected for a plasmid containing a small portion of the beta globin gene sequence that terminated between the Eco RI and Bam HI sites. That this was not the case was suggested by RNA-DNA hybridization tests with radioactive globin mRNA and by measurements of the size of the plasmid, which implied the presence of a large insert. Alternatively, the result suggested that plasmid pHb72 might carry an alpha globin gene sequence. The presence of an Eco RI site is consistent with

an alpha globin gene insert, since conventional sequence analysis of the globin mRNA's had already demonstrated that a single Eco RI site occurs in both the alpha and beta globin structural genes (9). In order to test this hypothesis, an Hae III restriction endonuclease digestion of plasmid pHb72 was carried out and the nucleotide sequence of one of the insert fragments was determined and compared with the known amino acid sequence of rabbit alpha globin (10).

To determine the number and size of the Hae III fragments that contain sequences from the gene insert, digests of pHb72 and the parent plasmid, pMB9, were electrophoresed in parallel on a 6 percent polyacrylamide gel. As expected, the pHb72 restriction pattern lacked an 850-base-pair pMB9 fragment containing the Eco RI site where the gene insertion was made. Three distinct new bands, whose sizes were estimated as 719, 438, and 113 base pairs, were observed, and their arrangement in the plasmid was ultimately determined to be as shown in Fig. 1. A band of unusual intensity suggested the presence of an additional "hidden" band about 120 base pairs in size, which comigrated with a pMB9 band. The existence of this band was subsequently confirmed by recovering it by Hae III digestion of the appropriate Alu I fragment.

The smallest of the insert bands was eluted from the acrylamide gel so that the "ladder" sequencing technique of Maxam and Gilbert (11) could be used to determine its nucleotide sequence. The fragment recovered from digestion of 100 μ g of pHb72 was treated with bacterial alkaline phosphatase. After this enzyme was inactivated by repeated phenol extractions, bacteriophage T4 polynucleotide kinase was used to transfer labeled phosphorus from [y-32P]adenosine triphosphate to the 5' termini of the fragment. The complementary strands were then separated by denaturation in 0.3NNaOH (for 30 minutes at 37°C) and electrophoresis on a tris-phosphate (0.036M tris, 0.03M NaH₂PO₄, pH 7.7) 4 percent polyacrylamide gel (bis-acrylamide: ac-





Fig. 2. (a) Autoradiograph of a ladder sequencing gel and the 47-nucleotide sequence derived from it. The 113-base-pair restriction fragment labeled at its Hae III terminal 5'-phosphate was strand-separated (see text). The fast strand was divided into six fractions, and each fraction was chemically cleaved at one of the four bases. The products were resolved by electrophoresis on a 20 percent acrylamide denaturing gel. The six tracks on the right are the same as the six on the left, except that they were electrophoresed into the gel for several hours longer. The base or bases cleaved as well as the relative amounts of cleavage in each reaction are indicated by the heavy letters at the top of each track. The reaction conditions used for each cleavage are given in (11) under the following descriptive titles: A > C, "alternative strong adenine, weak cytosine cleavage''; G, "alternative guanine reaction"; A > G, "strong adenine, weak guanine cleavage"; G > A, "strong guanine" weak adenine cleavage"; C, "cleavage at cytosine"; and C + T, "cleavage at thymine and cytosine." Brackets show the region of 15nucleotide overlap for the two experiments. Electrophoresis is from top to bottom. The heavy radioactive band at the bottom is the result of a trivial technical difficulty, a pileup of nucleotides due to a salt-front effect on the resolving gel. This artifact, which caused us to lose the first 21 nucleotides of sequence data at the 5' end of the fragment, can be avoided by running the gel beforehand for 4 hours at 400 volts. (b) A 77-nucleotide sequence derived by combining the ladder sequencing data with earlier sets of data that contributed to the sequence. The T1 fragment numbers refer to sequences reported in previous publications (9). The arrows show the directions and lengths of sequences that could be read from a ladder sequencing gel. Arrows running left to right were

read from the gel as shown. Arrows running right to left indicate that the complementary sequence was read from the gel. The data were obtained from an Hae III fragment with the labeled 5' ends 21 nucleotides in each direction beyond the portion shown here. Hha I, Alu I, and Hpa II sites exist in the sequence where indicated.

rylamide, 1:29 (11, 12). As estimated from autoradiograph intensity, about 80 percent of the material loaded on the gel was recovered as separated single strands. The rapidly and slowly migrating bands were eluted and divided into portions to be subjected to partial digestions specific for each of the four deoxynucleotides. In addition to the four standard reactions, we have found it useful to employ the "alternate A" and "alternate G" reactions developed by Maxam and Gilbert. The conditions of the reactions used have been described (7, 11).

The partial cleavage products were electrophoresed side by side on a tris-borate 20 percent polyacrylamide 7M urea gel. Figure 2a shows the results obtained for the rapidly migrating strand. The radioactive products from a particular cleavage all have common ³²P-labeled 5 ends and 3' ends that terminate at the nucleotide or nucleotides cleaved in the specific reaction. Separation according to size on the polyacrylamide gel enables the nucleotide sequence to be read directly from the autoradiograph, as indicated in Fig. 2a. This sequence is the complement of the mRNA sequence shown in Fig. 2b and corresponds exactly to the mRNA sequence coding for amino acid residues 87 to 103, as shown by the arrow pointing left under the sequence.

Since the sequence analyses presented here and the ladder sequence analyses we published earlier for the beta globin gene (7) correspond to known amino acid sequences, they provide an ideal opportunity to test the accuracy of the new sequencing approach developed by Maxam and Gilbert. Further tests of the accuracy are provided by comparison with the data we have obtained by more conventional sequencing techniques. In general, the accuracy of the method seems to be very good. In some cases a particular nucleotide may be difficult to read. The most troublesome example in Fig. 2a is the third nucleotide from the bottom of the left-hand set of gel tracks, that is, the third G in the sequence GGGAC (8). This corresponds to GUCCC in the complementary (message) strand. The only possible interpretations of the data at this point were G (bands in the G and G > Atracks) and C (bands of equal intensity in the C and C + T tracks). Since G residues frequently give bands of varying intensity in the C and C + T tracks, but C residues never give bands in the G track, it was felt that G was the only possible assignment. That this is the correct interpretation is shown by the fact that none of the other possibilities is consistent with the amino acid residue found at this position in the protein.

In addition to the data shown in Fig. 2a, we have obtained analogous sequencing data (not shown) from the complementary strand. From these it was possible to determine the sequence indicated by the arrow pointing right in Fig. 2b. In the large region where the arrows overlap, the nucleotide sequence has been determined entirely independently from the two sets of data, which are in complete agreement. Other checks on the accuracy of the sequence are provided by the ribonuclease T1 fragment sequences T1-54, T1-48, and T1-31 indicated in Fig. 2b. Again the correspondence is exact, indicating the accuracy of the sequencing technique as well as the fidelity of the maintenance of the cloned sequence in the Rec A⁺ host Escherichia coli C600. This particular sequence is remarkable in some respects, especially in having a content of 73 percent G plus C. The presence of an Hpa II site at amino acid residues 95-96 as well as another in a sequence reported earlier by Proudfoot and Brownlee (13) from the 3' untranslated region is also noteworthy. Since there are no such sites within the rabbit beta globin gene, digestion with Hpa II provides a convenient criterion for detecting plasmids carrying alpha globin sequences. Detailed discussion of the alpha globin sequences will be reserved for a subsequent report.

The National Institutes of Health guidelines for recombinant DNA research (14) specify that "when a cloned DNA recombinant has been rigorously characterized and there is sufficient evidence that it is free of harmful genes, then experiments involving this recombinant DNA can be carried out under P1 + EK1conditions if the inserted DNA is from a species that exchanges genes with E. coli and under P2 + EK1 conditions if not.' A footnote to the guidelines indicates that it is necessary to obtain approval from the NIH for the use of these reduced containment levels, such approval being "contingent upon data concerning: (a) the absence of potentially harmful genes (e.g., sequences contained in indigenous tumor viruses or which code for toxic substances), (b) the relation between the recovered and desired segments (e.g., hybridization and restriction endonuclease fragmentation analysis where applicable), and (c) maintenance of the biological properties of the vector.'

We have submitted data as specified above for the rabbit alpha globin plasmid pHb72, and on this basis the National Recombinant DNA Committee has approved its growth under P2 plus EK1 conditions. The rationale suggested in our petition is briefly summarized as follows. The method of joining the gene copy and

plasmid, homopolymer tailing, ensures that only one gene sequence can be inserted per plasmid. This was confirmed by measurements of the size of plasmid pHb72. When the gene copy is made from a monocistronic mRNA by using the cDNA cloning technique, it can be inferred with confidence that a globin gene copy cannot also contain other potentially harmful genes. Consequently, demonstration that the plasmid contains a desired globin gene sequence inserted by the cDNA cloning technique can help rule out the possibility that potentially harmful genes are present. The sequencing and restriction analyses presented here provide strong proof of the presence of the alpha globin sequence, which was corroborated by RNA-DNA hybridization data. Finally, the required proof of the "maintenance of the biological properties of the vector" includes demonstration of the maintenance of the tetracycline resistance gene, the ability of the plasmid to amplify in chloramphenicol, and comparison of pHb72 with the pMB9 parent plasmid by restriction fragmentation analysis with Hae III and other restriction endonucleases.

ALVIN Y. LIU, GARY V. PADDOCK HOWARD C. HEINDELL WINSTON SALSER

Department of Biology and Molecular Biology Institute, University of California, Los Angeles 90024

References and Notes

- 1. W. Salser, Annu. Rev. Biochem. 43, 923 (1974); see pp. 948-949
- A. Efstratiadis et al., Cell 7, 279 (1976).
 R. Higuchi, G. V. Paddock, R. Wall, W. Salser, Fed. Proc. Fed. Am. Soc. Exp. Biol. 35, 1369 (1976); Proc. Natl. Acad. Sci. U.S.A. 73, 3146 (1976); G. V. Paddock, R. Higuchi, R. Wall, W. Salser, Molecular Mechanisms in the Control of Gene Expression, D. P. Nierlich, W. J. Rutter, C. *F.* Fox, Eds. (Academic Press, New York, 1976), p. 535; W. Salser *et al.*, *Prog. Nucleic Acid Res. Mol. Biol.* **19**, 177 (1976). The cloning and all work with live cells carrying globin sequences were carried out under P3 conditions under Asilomar guidelines before EK2 systems were certified. Similar experiments are now being carried out under P3 conditions with the newly certified EK2 systems, as is now specified. T. Maniatis, S. G. Kee, A. Efstratiadis, F. C.
- Manaus, S. G. Ree, M. Damanne, M. Kafatos, Cell 8, 163 (1976).
 B. W. O'Malley, S. L. C. Woo, J. J. Monahan, L. McReynolds, S. E. Harris, M. J. Tsai, S. Y. Tsai, A. R. Means, in Molecular Mechanisms in the Conference of D. P. Nierlich W. J. 5. A. R. Means, in Molecular Mechanisms in the Control of Gene Expression, D. P. Nierlich, W. J. Rutter, C. F. Fox, Eds. (Academic Press, New York, 1976), p. 309; F. Rougeon and B. Mach, Proc. Natl. Acad. Sci. U.S.A. 73, 3418 (1976). F. Rougeon, P. Kourilski, B. Mach, Nucleic Acids Res. 2, 2365 (1975); T. H. Rabbits, Nature (London) 260, 221 (1976); K. O. Wood and J. C. Lee, Nucleic Acids Res. 3, 1961 (1976).
- J. Browne, G. V. Paddock, A. Liu, P. Clarke, H. Heindell, W. Salser, *Science* **195**, 389 (1977).
- Abbreviations for the amino acid residues are Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Glu, glutamic acid; Gln, glutamine; Gly, glycine; His, histidine; 8. Ile, isoleucine; Leu, leucine; Lys, lysine; Met. methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine. Other abbrevi-ations are A, adenine; C, cytosine; U, uracil; G, guanine; and T, thymine. W. Salser *et al.*, *Fed. Proc. Fed. Am. Soc. Exp.*
- 9. W. Salser et al., Fed. Proc. Fed. Am. Soc. Exp. Biol. 35, 23 (1976); G. V. Paddock, R. Poon, H

SCIENCE, VOL. 196

Heindell, J. Isaacson, W. Salser, J. Biol. Chem.,

- in press. 10. M. O. Dayhoff, Atlas of Protein Sequence and M. O. Dayhoff, Atlas of Protein Sequence and Structure (National Biomedical Research Foun-dation, Silver Spring, Md., 1972), vol. 5.
 A. M. Maxam and W. Gilbert, Proc. Natl. Acad. Sci. U.S.A. 74, 560 (1977).
 G. S. Hayward, Virology 49, 342 (1972).
 N. J. Proudfoot and G. G. Brownlee, Nature (London) 263, 211 (1976).
 Dependence of Hough Education and Welfare
- 11.
- 12.
- Department of Health, Education and Welfare, Fed. Regist. **41** (No. 131), 27917 (1976). 14.
- 15. We thank R. Higuchi for his help during the early stages of this work and A. Maxam and W. Gilbert for supplying details of their technique in adfor supplying details of their technique in ad-vance of publication. This research was support-ed in part by PHS grants GM 18586 and CA 15940 and by grant VC-240 from the American Cancer Society. G.V.P. was supported by a Helen Hay Whitney Fellowship. A.L. and H.C.H. were sup-ported by National Cancer Institute predoctoral training grant 783825-32923-7.

7 February 1977

Analysis of Chicken Ribosomal RNA Genes and Construction of Lambda Hybrids Containing Gene Fragments

Abstract. The ribosomal RNA genes (ribosomal DNA) of chicken are present in approximately 200 copies and are cleaved into two fragments of molecular weight 5 imes 10^6 and 12 to 14×10^6 by restriction endonuclease Eco RI. Recombinant phages have been constructed in vitro by joining the smaller fragment of ribosomal DNA and the outer arms of DNA from the vector $\lambda gt WES \lambda C$. In one of the recombinants, the coding strand of the cloned fragment is in the proper orientation for transcription with λ "early" genes; in the other two the orientation is reversed, with the coding strand in the proper position for transcription with λ 'late' genes.

Cloning of ribosomal RNA (rRNA) genes, which are present in multiple copies and specify a readily available RNA probe, can provide a convenient system for detailed analysis of eukaryotic DNA (1). In addition to yielding hybrid DNA which includes genes that are interesting in themselves, these experiments provide an opportunity to develop methods potentially useful for the cloning of unique sequence genes (2). We have employed the certified EK2 vector $\lambda gtWES \cdot \lambda C$ (3) to clone and characterize a portion of the ribosomal gene from chickens (Gallus domesticus).

Ribosomal RNA genes represent 0.02 to 0.12 percent of chicken DNA (4, 5) and are clustered. If these genes are in tandem, as are ribosomal DNA (rDNA) sequences in other eukaryotic chromosomes (6-8), then cleavage with restriction endonuclease should yield a number of identical fragments. Figure 1 shows results from experiments in which chicken





Fig. 1. Agarose gel analysis of chicken DNA and recombinant phage DNA. Highmolecular-weight DNA was prepared from erythrocytes of Spafas (Norwich, Conn.) gs⁻ chickens by a method similar to that of Chambon and co-workers (18). The DNA was incubated in 0.1M tris (pH 7.4), 0.05M NaCl, 0.005M MgCl₂, and 0.001M p-chloromercuribenzoate for 10 minutes at 37°C. Eco RI was added (2 units per microgram of DNA) and the mixture was incubated for 10 minutes. A second portion of Eco RI was then added and incubation was continued for 10 minutes. The reaction was stopped by adding one-tenth volume 10 percent sodium dodecyl sulfate and heating to 65°C for 2 to 5 minutes. Digested DNA was concentrated by ethanol precipitation before electrophoresis; such treatment causes some loss of large fragments. A 10- to 20-µg portion of the concentrated DNA was subjected to electrophoresis on 0.7 percent agarose disc gels (0.5 by 12 cm) in buffer containing 0.04M tris, 0.005M sodium acetate, and 0.001M EDTA (the pH was adjusted to 7.9 with acetic acid). Gels were stained for 1 hour or more in ethidium bromide (0.5 μ g/ml). The DNA was transferred to cellulose nitrate membranes by the method of Southern (8). The 18S and 28S rRNA probes (specific activity $\simeq 4 \times 10^6$ cpm/µg) were prepared from chicken fibroblasts labeled with [3H]uridine. Each probe was shown to be homogeneous by polyacrylamide gel electrophoresis. The concentration of probe RNA in the hybridization mixture was approximately 0.5 μ g/ml. Hybridization was detected by fluorography (8, 19). The stoichiometry of the recombinant DNA Eco RI fragments was determined from densitometer tracings of a photographic negative of ethidium bromide-stained gels. The area under each peak was determined by cutting out and

weighing that peak. Molecular weights were taken to be 13.8×10^6 and 9.0×10^6 for the vector arms and 5.0×10^6 for the inserted fragment. Normalizing to the smaller vector fragment, the ratios of fragments (from largest to smallest) are: λ Gd1, 0.98 : 1.00 : 0.76; λ Gd2, 1.00 : 1.00 : 0.91; and λ Gd3, 1.08: 1.00: 0.86. Lambda Gd1 DNA, uniformly labeled with ³²P, was digested with Eco RI and subjected to electrophoresis on 0.7 percent agarose gels. The radioactivity in each fragment was determined by cutting out the ethidium bromide-stained bands and counting them in 5 ml of Aquasol. The counts were (from largest to smallest) 297, 154, and 80 cpm. Normalizing to the smaller vector arm, the ratio of fragments is 1.25 : 1.00 : 0.94. (a) Ethidium bromide-stained gel containing Eco RI fragments of unfractionated chicken DNA. (b) Fluorogram of Southern transfer containing Eco RI fragments of unfractionated chicken DNA hybridized with 18S plus 28S 3H-labeled rRNA. In this transfer, hybridization to the 5 \times 10⁶ dalton fragment is apparent but hybridization to the 12 to 14 \times 10⁶ dalton fragment is not. (c) Fluorogram as in (b) except that before digestion with Eco RI, DNA was enriched three- to fivefold for ribosomal sequences by selecting material from the heavy side of DNA banded in a CsCl equilibrium gradient. In this case hybridization to both fragments is evident. (d to f) Ethidium bromide-stained gels of Eco RI-digested DNA from (d) λ Gd1, (e) λ Gd2, and (f) λ Gd3.