Reports

Nucleotide Sequences from the Rabbit Beta Globin Gene Inserted into *Escherichia coli* Plasmids

Abstract. *A 169-nucleotide region from the rabbit beta globin gene has been sequenced by analysis of complementary DNA's cloned in bacterial plasmids. Comparison of these sequences with those established for this gene by other techniques provides evidence of a high degree of fidelity and allows the unambiguous establishment that these plasmids do not contain harmful sequences.*

We have recently cloned rabbit beta globin complementary DNA (cDNA) in a bacterial plasmid (*1*). Cloning provides us with the large quantitites of duplex beta globin gene sequences needed in order to use the new rapid "ladder" sequencing technique developed by Maxam and Gilbert (*2*). As a first step toward completely sequencing the rabbit beta globin gene, we have sequenced several regions that span the interval coding for amino acid residues 93 through 146 and into the untranslated region at the 3′ end of the messenger RNA (mRNA). The sequence obtained shows only one discrepancy with the known amino acid sequence (*3*) and with sequences previously determined directly from rabbit beta globin mRNA and cDNA (*4–6*). The single discrepancy noted results from the elimination of a single nucleotide in the cloned sequence. These results establish that the maintenance of fidelity of sequences inserted into a bacterial plasmid under these conditions is good but not perfect.

Combining the "ladder" sequencing technique with cDNA cloning techniques greatly improves both the ease and confidence with which mammalian mRNA's can be sequenced. In fact, analysis of a significantly long region of a sequence inserted in a bacterial plasmid has become so rapid and easy that the ladder sequencing technique is a useful assay to unambiguously characterize a recombinant DNA molecule.

The starting material for the DNA sequence analysis shown in Fig. 1 was the bacterial plasmid pHb23, which had been shown to contain the rabbit beta gene sequences by two-dimensional electrophoretic analysis of enzymatic digests of mRNA sequences which annealed to the plasmid (*1*). Comparison of the restriction endonuclease Hae III fragment pattern of pHb23 with that of its parent plasmid pSC101, by electrophoresis on an 8 percent acrylamide slab gel, shows that pHb23 has two high-molecular-weight fragments not found in the pSC101 pattern. These new fragments must carry portions of the inserted sequence. As expected, the pSC101 pattern contains a low-molecular-weight band that is not present in the recombinant plasmid and has been shown to contain the Eco RI site used for the insertion of the globin gene sequence.

A double digest with restriction endonuclease Eco RI and Hae III shows that the higher molecular weight insert band is cleaved by Eco RI. Previous sequencing work (*4*, *5*) had shown an Eco RI site within the structural gene at amino acid residues 120–121. The results reported below confirm that the RI site seen is at amino acid residue 120–121 and rule out the alternative possibility that the site seen could actually be from one of the untranslated regions. Of the two fragments resulting from the Eco RI cleavage, the smallest was of the order of 40 to 60 nucleotides in length as judged by acrylamide gel electrophoresis. Examination of the amino acid sequence shows that the only possible Hae III sites in this region which could give fragments in this size range are at amino acid residue 142, which is 63 nucleotides from the probable RI site, and amino acid residue 138, which is 51 nucleotides from the probable RI site.

We have sequenced this fragment by the "ladder" sequencing technique (*2*) as described below. Hae III restriction endonuclease was used to cleave 50 μg of pHb23. The products were separated by electrophoresis on an 8 percent acrylamide slab gel. The largest insert band was eluted as described (*7*), except that no carrier transfer RNA (tRNA) was used. The restriction fragment was then labeled at its 5′ terminal phosphate by first digesting with bacterial alkaline

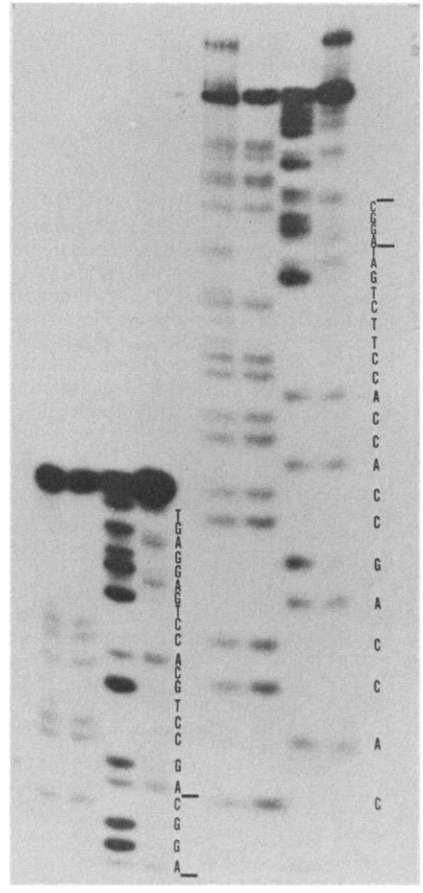

Fig. 1. An autoradiograph of a ladder sequencing gel and the 43-nucleotide sequence derived from it. The 51-nucleotide restriction fragment, labeled at its Hae III terminal 5′ phosphate, was divided into four fractions and each fraction was chemically cleaved at one of the four bases. The products were separated by electrophoresis on a 20 percent acrylamide denaturing gel: The four tracks on the left are the same as the four on the right, except that they migrated into the gel several hours before the second set was layered onto the gel. Lanes 1 and 4 are the result of a chemical cleavage where cleavage occurs at cytidine more frequently than at thymidine. In lanes 2 and 5 cleavage is only at cytidine. In lanes 3 and 7 cleavage occurs at guanosine more frequently than at adenosine. In lanes 4 and 8 cleavage occurs at adenosine more frequently than at guanosine residues. Brackets show the region of overlap. Electrophoresis is from top to bottom.

phosphatase, leaving 5′ hydroxyl groups, and then treating with bacteriophage T4 polynucleotide kinase and γ-^{32}P–labeled adenosine triphosphate. The fragment was then cleaved with a second restriction enzyme and acrylamide gel electrophoresis was used to separate the two new labeled fragments. A labeled fragment was then excised from the gel, eluted, and divided into four fractions. Each fraction was chemically cleaved more or less preferentially at one of the four bases. Only a mild treatment was used, sufficient to give fragments which average about 40 nucleotides in length. Thus, when preferentially cleaving at guanosine residues, one obtains labeled fragments corresponding to all lengths that reach from the labeled 5′ terminus to guanosine residues. These were then separated by electrophoresis on a 20 percent acrylamide denaturing gel. When all four reactions are run side by side, the sequence of the fragment can easily be read (*8*).

Figure 1 shows the autoradiograph of a sequencing gel and the 43 nucleotide sequence derived from it. In this experiment, we obtained the sequences complementary to the beta globin mRNA. In other experiments, by reversing the procedure so that the first digestion was with Eco RI, followed by labeling the 5′ termini and then digesting with Hae III, we were able to obtain corresponding data from the strand with the sequence of the mRNA itself (not shown). Combining the sequence shown with sequences of the restriction endonuclease sites, the sequence of the 51 nucleotide restriction fragment extending from amino acid residue 137 (Val) to residue 123 (Thr) is completed (*9*).

We have carried out similar experiments with other restriction fragments from this and other recombinant DNA's. Figure 2 shows the nucleotide and amino acid sequence obtained in this way for rabbit beta globin from amino acid 93 (Cys) to four nucleotides beyond the termination codon. (Brackets under the sequence indicate the different sets of data that contributed to this 169 nucleotide sequence.) The ribonuclease T1 fragment numbers correspond to those used earlier (*4, 6*). There are three ambiguities left in this sequence. The ambiguity at amino acid residue 142 (Ala) is especially interesting. When the sequence from the ladder gel extending from a Hae III site at amino acid residue 138 through this site (which clearly shows a continuous sequence) is fit to the amino acid sequence, it becomes apparent that the third position nucleotide of 142 (Ala) or one of the bordering nucleotides has been deleted. We cannot say whether this deletion occurred while making the cDNA from the mRNA, while making the cDNA double stranded, or subsequently during replication of the cloned sequence. A fourth alternative, that the mRNA sequence, which was cloned, itself contained the deletion, seems less likely. So far as we know, this is the most complete published assessment of the fidelity of the replication of cloned sequences at the single nucleotide level. Growth of plasmid pHb13 was in the Rec A$^-$ host HB101.

Since a single experiment can yield a long sequence (for example, the 43-nucleotide sequence shown in Fig. 1) this approach provides a convenient way to rapidly and unambiguously characterize recombinant DNA molecules carrying globin gene sequences. pHb23 and all other recombinant molecules described in this report were made from mRNA's by means of a cDNA cloning technique (*1*). The method of joining, homopolymer tailing (*1*), ensures that only one gene sequence can be inserted per plasmid.

In theory, two or more inserts could become associated in alternation with concatenated plasmids, but this has never been observed and would be detected since electrophoresis of all purified prep-

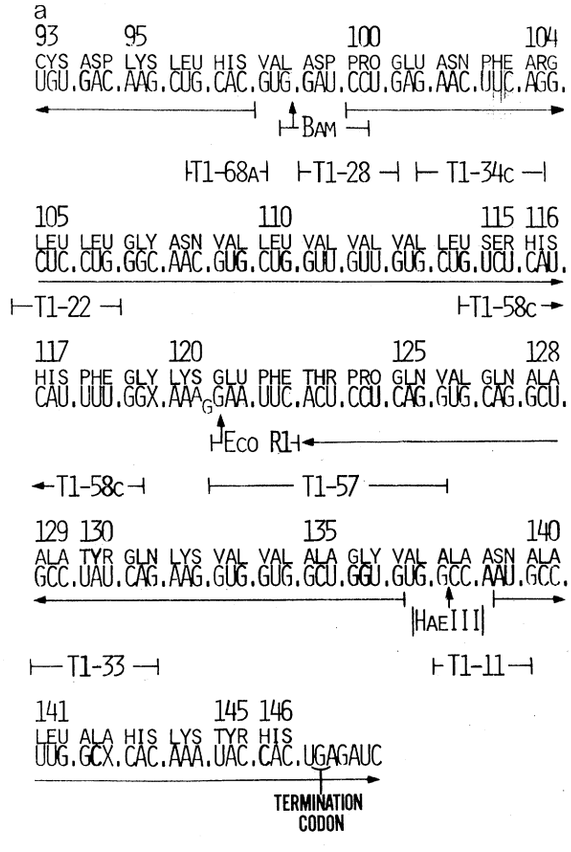

Fig. 2. The 169-nucleotide sequence derived by combining the ladder sequencing data with previously determined sequences. The amino acids and the different sets of data that contributed to the sequence are also shown. The T1 fragment numbers refer to sequences determined previously (*4–6*). The arrows show the direction and length of a sequence that could be read on a ladder sequencing gel. Arrows running left to right were read as shown. Arrows running right to left indicate that the complementary sequence was read from the gel. Some sequences were determined from the site specificity of the various restriction endonucleases as indicated. The two ladder sequences extending from the Bam restriction site cleavage between the codons at 98–99 used restriction fragments from pHb56, which does not extend toward the 5′ end further than amino acid position 93. After the 5′ ends at the Bam site were labeled with T4 polynucleotide kinase, a second cleavage was carried out with Hae III. The region between the Eco RI site at amino acid position 121 and the Hae III site at amino acid position 138 was determined with the use of pHb23 as described in the text. The region extending from the Hae III site at amino acid position 138 past the termination codon was determined using a Hae III restriction fragment from pHb13. After the 5′ ends of this fragment were labeled, the strands were denatured and separated on a 4 percent acrylamide gel (the ratio of bisacrylamide to acrylamide was 1 : 29) prior to ladder sequencing. The X in the sequence at the 142 (Ala) indicates a nucleotide which is deleted in plasmid pHb13.

arations was carried out with standards on agarose gels. It is known that alpha and beta globins are coded by monocistronic mRNA's. The mRNA used for cloning was cytoplasmic in origin (obtained by release from purified polysomes), and selected for size. Therefore, it can be inferred with confidence that it does not contain any hazardous gene sequence. By contrast, a clone made from nuclear DNA or RNA and shown to contain the beta globin sequence might still be hazardous since it could contain neighboring nucleotide sequences which might include an unknown hazardous gene sequence.

Even considering the degeneracy of the genetic code, the chance that a nucleotide sequence 30 or more nucleotides in length will have a fortuitous fit with an amino acid sequence the size of the globin chains is less than 1 in 10^9. Consequently, when inserts are made by the cDNA technique, the amino acid sequence is known, the mRNA is known to be monocistronic, and the known protein product is deemed harmless, then determination of a sequence of 30 or more nucleotides within the structural gene can provide an effective approach for establishing that a plasmid does not contain a harmful sequence.

We have used this approach and other characterization data as the basis for petitioning for the reduction of the biocontainment requirements from P3 plus EK2 to P2 plus EK1 for pHb23 and other alpha and beta globin plasmids derived from rabbit and chicken. This is the procedure specified in the recombinant DNA guidelines (*10*) for those clones of DNA experiments recombinants derived from shotgun (cloning with DNA preparations less than 99 percent pure) which have been thus characterized.

Thus far we have emphasized the use of our cloned mRNA sequences as a means for determining mRNA sequences. Now that several of these recombinant globin plasmids have been extensively characterized we can use them as hybridization probes for gene enrichment to permit cloning of chromosomal DNA fragments containing the very interesting regions adjacent to the globin structural genes.

JEFFREY K. BROWNE
GARY V. PADDOCK
ALVIN LIU, PAT CLARKE
HOWARD C. HEINDELL
WINSTON SALSER

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

28 JANUARY 1977

References and Notes

1. R. Higuchi, G. V. Paddock, R. Wall, W. Salser, *Proc. Natl. Acad. Sci. USA* **73**, 3146 (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.
2. A. Maxam and W. Gilbert, personal communication, and *Proc. Natl. Acad. Sci. U.S.A.*, in press.
3. M. O. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md., 1969).
4. R. Poon, G. V. Paddock, H. Heindell, P. Whitcome, W. Salser, D. Kacian, A. Bank, R. Gambino, F. Ramirez, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3502 (1974).
5. W. Salser, S. Bowen, D. Browne, F. El Adli, N. Fedoroff, K. Fry, H. Heindell, G. V. Paddock, R. Poon, B. Wallace, P. Whitcome, *Fed. Proc.* **35**, 23 (1976).
6. W. Salser, J. Browne, P. Clarke, H. Heindell, R. Higuchi, G. V. Paddock, J. Roberts, G. Studnicka, P. Zakar, *Prog. Nucleic Acid Res. Mol. Biol.* **19**, 177 (1976).
7. W. Gilbert and A. Maxam, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3581 (1973).
8. Thymidine- and cytidine- specific cleavages are begun by depyrimidization in 15*M* hydrazine (Eastman) with 20 μg of salmon sperm carrier DNA for 50 minutes at 20°C. Cytidine specific depyrimidization is done with 15*M* hydrazine, 1*M* NaCl, and 20 μg of salmon sperm DNA for 100 minutes at 20°C. Strand scission was accomplished by heating the samples for 30 minutes at 90°C in 0.5*M* aqueous piperidine. The samples were then dried and suspended in 0.1*M* NaOH. An equal volume of a solution containing 10*M* urea, 0.02 percent xylene cyanol FF, and 0.02 percent bromophenol blue was added to each sample. Purine specific cleavage was achieved by first methylating the bases by incubating samples in a solution of 50 m*M* sodium cacodylate (*p*H 8.0), 10 m*M* $MgCl_2$, 0.1 m*M* EDTA, 20 μg of salmon sperm DNA, and a 1/250 dilution of dimethyl sulfate (Eastman) for 60 minutes at 20°C. The reaction was stopped by the additon of a 1/10 volume of a solution containing 2.5*M* 2-mercaptoethanol, 3*M* sodium acetate (*p*H 7.5), 0.1*M* magnesium acetate, and tRNA (1 mg/ml). Base release of guanine residues was accomplished by heating at 90°C for 15 minutes. Base release of adenine residues was carried out at 0°C for 60 minutes in 0.2*N* HCl. Strand scission for the purines was accomplished by heating for 15 minutes at 90°C at an alkaline *p*H. The samples were then mixed with an equal volume of the urea dye solution. One-half of each fraction was layered on a gel (0.2 by 20 by 40 cm) consisting of 20 percent acrylamide, 0.67 percent bisacrylamide, and 7*M* urea. The gel buffer and the running buffer consisted of 50 m*M* trisborate (*p*H 8.3) and 1 m*M* EDTA. The samples were placed on the gel and subjected to electrophoresis until the bromophenol blue dye marker had migrated 36 cm. The second half of each sample was then layered on in a second set of tracks, and electrophoresis was continued until the bromophenol blue dye marker of the second set had migrated approximately 22 to 23 cm. The gel was visualized by autoradiography.
9. 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; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Try, tryptophan; Tyr, tyrosine; and Val, valine. Other abbreviations are A, adenine; C, cytosine; U, uracil; and G, guanine.
10. Department of Health, Education, and Welfare, National Institutes of Health, "Recombinant DNA research," *Fed. Reg.* **41** (No. 131), 27917 (1976).
11. We thank A. Maxam and W. Gilbert for supplying details of their technique in advance of publication and we thank S. Weissman and B. Forget for the sequences they determined for human globin mRNA. This research was supported in part by PHS grants GM18586 and CA15940. J.B. was supported by PHS molecular biology training grant GM-1351. G.V.P. was supported by a Helen Hay Whitney Fellowship. A.L. and H.C.H. were supported by National Cancer Institute predoctoral training grant 4-783825-32923-7.

13 December 1976

Novel Screening Procedure for Recombinant Plasmids

Abstract. *Lysed bacterial colonies containing potential recombinant plasmids were mixed with molten agar and sealed into slots of an agarose gel. After electrophoresis overnight, the gel was stained with ethidium bromide, which clearly reveals recombinant plasmids.* Xenopus laevis *ribosomal DNA and histone DNA of* Psammechinus miliaris *were ligated into pCRI plasmids and screened by this method.*

DNA sequences of particular interest are often recognized to be associated with restriction fragments of defined length (*1*). In these cases, the construction of recombinants is greatly facilitated provided that screening methods are employed which depend on recognizing particular DNA size classes.

Histone DNA (hDNA) was purified from sperm DNA of *Psammechinus miliaris* by actinomycin C_1/CsCl centrifugation (*2*). The DNA was then restricted with HindIII restriction endonuclease, which cleaves the hDNA cluster once every 6 kb (1 kb = 10^3 base pairs) repeat (*1*). The plasmid pCRI has only one HindIII site, which resides in the gene or control region conferring kanamycin resistance (*3, 4*). Thus, integration of foreign DNA at this site prevents expression of kanamycin resistance, but the resistance to colicin is unaffected. This property was used to distinguish between ligated pCRI molecules and pCRI recombinants. The 6-kb fragments of cellular histone DNA or fragments reclaimed from the recombinant λSam7h22 (*5*) were ligated with HindIII restricted pCRI, and the products were used to transfect $CaCl_2$-treated HB101. Selection was performed on L plates containing colicin EI (*6*). Colicin-resistant colonies were picked onto L plates containing kanamycin (100 μg/ml), and colonies unable to grow were tested by the method described below.

Xenopus laevis ribosomal DNA (rDNA) was prepared from total blood DNA (*7*) by two runs on CsCl gradients containing actinomycin (*2*) and after removal of the actinomycin on a third normal CsCl gradient. The peak fractions