method described here provide a source from which we are able to obtain relatively large quantities of individual repetitive DNA sequences. Such sequences can be used to isolate all the related members of given repetitive sequence families from the genome. The method described can, of course, be used for cloning any DNA sequence without the requirement that it be terminated by specific restriction enzyme sites (*6*), and without the addition of substantial homopolymer sequences to the cloned fragment.

RICHARD H. SCHELLER
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena 91125
TERRY L. THOMAS, AMY S. LEE
WILLIAM H. KLEIN, WALTER D. NILES
ROY J. BRITTEN, ERIC H. DAVIDSON
Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

1. This work was done in compliance with the NIH Guidelines for Recombinant DNA Research.
2. D. E. Graham, B. R. Neufeld, E. H. Davidson, R. J. Britten, *Cell* **1**, 127 (1974).
3. F. C. Eden, D. E. Graham, E. H. Davidson, R. J. Britten, *Nucleic Acids Res.*, in press.
4. T. Ando, *Biochim. Biophys. Acta* **114**, 158 (1966).
5. R. H. Scheller, R. E. Dickerson, H. W. Boyer, A. D. Riggs, K. Itakura, *Science* **196**, 177 (1977).
6. V. Sgaramella, J. H. Van de Sande, H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1468 (1970); H. L. Heyneker, J. Shine, H. M. Goodman, H. W. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S-Y. Lin, A. D. Riggs, *Nature (London)* **263**, 748 (1976); C. P. Bohl *et al.*, *Gene* **1**, 81 (1977).
7. M. So, R. Gill, S. Falkow, *Mol. Gen. Genet.* **142**, 239 (1975).
8. S. N. Cohen, C. Y. Chang, L. Hsu, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2110 (1972).
9. P. Fredericq, *Annu. Rev. Microbiol.* **11**, 7 (1957).
10. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
11. T4 polynucleotide kinase, T4 DNA ligase, and Eco RI endonuclease were gifts of P. Green, H. Heyneker, and H. Boyer. Supported by NIH grants HD-05753 and GM-20927 and by NSF grant BMS 75-07359. R.H.S. is supported by an NIH predoctoral training grant; T.L.T. by an NIH postdoctoral fellowship; A.S.L. by fellowship J-289 and W.H.K. by Lievre fellowship J-340, both from the American Cancer Society, California Division.

7 February 1977

Human Globin Messenger RNA: Importance of Cloning for Structural Analysis

Abstract. *The sequence of most of the human beta globin messenger RNA and large sections of the alpha globin messenger RNA has been determined. Partly because of genetic polymorphism, it was necessary to clone globin complementary DNA in order to extend the analysis. Purified human fetal globin messenger RNA was isolated and used as a template by reverse transcriptase to produce duplex complementary DNA molecules. These molecules were linked in vitro to plasmid DNA by use of T4 ligase in the presence of* Escherichia coli *Pol 1. Several colonies transformed by these molecules have been shown to hybridize with labeled human globin complementary RNA.*

By use of various RNA and DNA sequencing methods, we recently completed the nucleotide sequence of almost the entire translated and 3′ untranslated region of human beta globin messenger RNA (mRNA) (*1–3*). We are currently sequencing both the 3′ untranslated and the translated sequence of human alpha globin mRNA. To complete the analysis we found it necessary to proceed by use of gene cloning procedures, because base sequence analysis of an individual's globin mRNA may show the existence of genetic polymorphism. With the presence of heterogeneities, one cannot join specific base sequences in different regions. Cloning techniques permit one to obtain the large amounts of purified single-species DNA necessary for sequencing analysis.

Most of the mutations resulting in single amino acid substitutions that have been studied in human hemoglobin (Hb) can be accounted for by a single base change in the DNA. However, in beta globin mRNA's there are mutations that cannot be accounted for by such single base substitutions in a unique normal "ancestor" sequence. There are three amino acid codons in the beta globin mRNA for which protein or nucleic acid studies provide evidence for genetic polymorphism. One occurs at amino acid position 20. Normal globin has valine at this position [codon GUN (*4*)]. From RNA sequencing procedures described elsewhere (*5*), we have determined the nucleotide sequence of at least some of the molecules to be GUG (*1*). Amino acid substitutions for amino acid 20 in the Hb molecule have been found in the general population. One such substitution, Hb Olympia, has methionine (AUG) at position 20. Another, Hb Strasbourg, has aspartate (GAY) (*6*).

Table 1. Results of transformation experiments.

DNA species	Transformants after cycloserine selection
No DNA	0
Native PBR313 DNA	0
PBR313 DNA + Hb cDNA	0
PBR313 DNA + Hb cDNA + ligase	60

A second example of potential polymorphism in beta globin mRNA comes from amino acid substitution data at amino acid 67. Normal hemoglobin mRNA codes for valine (GUN); however, Hb Bristol mRNA codes for aspartate (GAY) and Hb M Milwaukee mRNA codes for glutamate (GAR). The third position of the codon for amino acid 67 could therefore be U or C as well as A or G. In our sequencing data, we find only GUG (*1–3*).

The third example of polymorphism in beta globin mRNA exists at codon position 50 (ACN), which codes for threonine. Here cDNA sequencing analysis from one patient indicated the codon to be ACU, whereas complementary RNA (cRNA) analyses of mRNA from another patient (*3*) were consistent with ACA. The amino acid substitution in Hb Edmonton ($Thr^{50} \rightarrow Lys$) suggests the presence of a purine (A or G) as the third base of this codon. These data also support the presence of genetic polymorphism in the third nucleotide position.

Since the alpha chain is known to be coded by duplicate loci per haploid set of chromosomes in a substantial portion of the population (*6*), one would expect more polymorphism in alpha mRNA than in beta mRNA. One such example is present in the untranslated region near the 3′ terminus. Partial venom diesterase digestion of a cDNA fragment from this region showed an interchange between two bases in approximately half the molecules (Fig. 1). We derived the following oligonucleotide sequence from these tracks: either CCCCUCCU or CCCUCCCU. The interchanged nucleotides are residues 60 and 61 beyond the termination codon (unpublished results). This heterogeneity has been found in at least two patients.

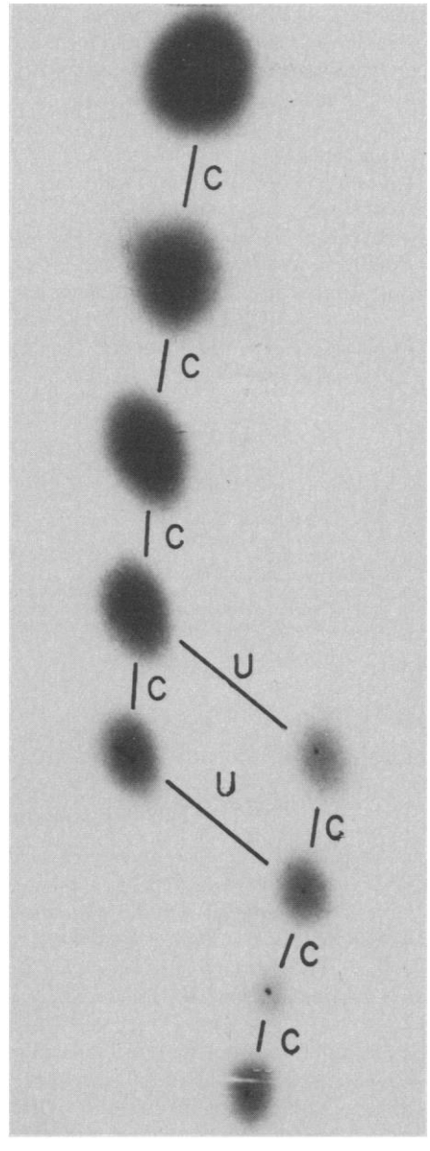
In addition to resolving the problems created by polymorphism, cloning procedures would make practical a number of other sequencing projects. For example, the 5′ terminal residues are not well represented in our current cDNA preparations and are therefore difficult to sequence. Cloning is also necessary to obtain sufficient amounts of purified minor hemoglobin species such as HbF (γ chain) and HbA_2 (δ chain), and for the deficient β mRNA in β thalassemia.

With the issuance of the Asilomar guidelines at the end of the voluntary moratorium, we decided to clone highly purified fetal globin cDNA, using procedures similar to those used successfully by others to clone rabbit globin cDNA (*7, 8*) RNA of fetal red blood cells, from a newborn infant undergoing exchange transfusion, was purified by sucrose gradient centrifugation followed by oligo(dT) cellulose column chromatography and gel electrophoresis. Globin mRNA purified by this technique contains no detectable RNA species other than 10*S* globin mRNA when analyzed by polyacrylamide gel electrophoresis and nucleotide sequencing techniques (*2*).

Fetal human globin mRNA purified as described above was used as template for the synthesis of a synthetic DNA copy of the globin genes by a two-step procedure involving viral reverse transcriptase: (i) synthesis of a single-stranded cDNA copy of the mRNA in the presence of actinomycin D (*9*) (Fig. 2A); and (ii) after alkaline hydrolysis of the mRNA and isolation of the cDNA by Sephadex column chromatography, synthesis of a second strand complementary to the cDNA in a self-primed reaction (*8, 10*), using essentially the same conditions as those used in the first synthetic step but without actinomycin D, oligo(dT), or mRNA (Fig. 2B). The second step resulted in the conversion of approximately 30 percent of the "full length" single-stranded cDNA into DNA with an electrophoretic mobility significantly lower than that of the initial product and consistent with a "hairpin" DNA molecule of the size of a double-stranded transcript of the globin mRNA (Fig. 2B) (*10*).

The resistance of the double-stranded cDNA obtained in step (ii) to degradation by S1 nuclease was 66 percent. Because of the evidence that such a cDNA is a hairpin structure (*8, 10*), the double-stranded cDNA was then digested with S1 nuclease to degrade the hairpin and provide an "open" duplex DNA molecule suitable for joining to a plasmid. After digestion with S1 nuclease and polyacrylamide gel electrophoresis in the presence of 98 percent formamide, the cDNA preparation was found to migrate as a prominent peak with a mobility approximately equal to that of the original single-stranded cDNA; a number of less prominent and more rapidly migrating peaks were also present (data not shown).

For insertion of the fetal globin cDNA (HbF cDNA) into plasmid DNA, the DNA of plasmid pMB9 or pBR313 was cleaved with the restriction endonuclease



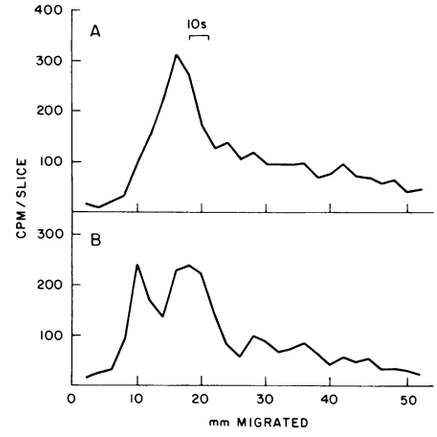


Fig. 1 (left). Autoradiographs of two-dimensional fractionation of partial snake venom phosphodiesterase (SVP) digest of a fragment of Hb cDNA. The fragment labeled at the 5′ terminus was digested with SVP (*3*) and fractionated by electrophoresis from left to right at *p*H 3.5 and homochromatography from the bottom upward as described in (*3*). The oligonucleotide map shows the possibility of two sequences: CCCUCCC and CCCCUCC.

Fig. 2 (above). Polyacrylamide gel electrophoresis of human globin cDNA used in cloning experiments. Polyacrylamide gel electrophoresis was carried out in the presence of 98 percent formamide as previously described (*9*). The disc gels were sliced at 2-mm intervals, and the radioactivity of each slice was determined as described (*9*). The bracket indicates the position of 10*S* globin mRNA run on a parallel gel and detected by staining with methylene blue. (A) Single-stranded cDNA initially synthesized from globin mRNA in the presence of [^{3}H]deoxycytidine triphosphate and actinomycin D. (B) Double-stranded cDNA obtained by incubating the single-stranded [^{3}H]cDNA in (A) with avian myeloblastosis virus DNA polymerase, using only nonradioactive deoxynucleotide triphosphates in a self-primed reaction without the addition of mRNA, oligo(dT), or actinomycin D; *CPM*, counts per minute.

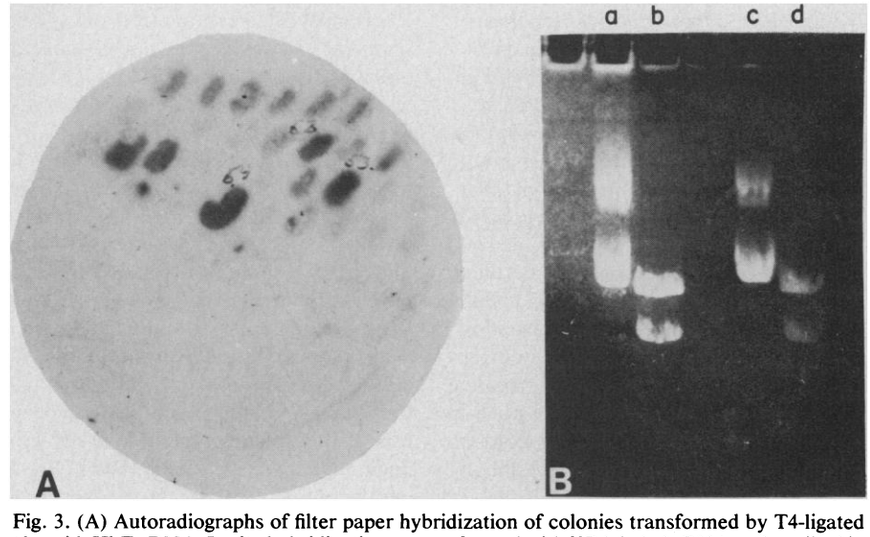


Fig. 3. (A) Autoradiographs of filter paper hybridization of colonies transformed by T4-ligated plasmid–HbF cDNA. In situ hybridization was performed with ^{32}P-labeled cRNA as described in the text. Dark areas represent transformed colonies hybridizing to Hb cRNA. (B) Electrophoresis of potential hybrid DNA recombinants. Columns a and c represent plasmid DNA (form I and form II) isolated from transformed colonies; columns b and d, pBR313 plasmid DNA (form I and form II).

Eco RI or Bam, respectively. The linear plasmid DNA was mixed with double-stranded HbF cDNA and the mixture was treated with S1 nuclease, followed by T4 ligase and *Escherichia coli* DNA polymerase I. The ligated DNA's were used to transform *E. coli* strain HB101. We then looked for colonies containing hybrid DNA molecules by selecting for antibiotic resistance (pMB9) or, as in the case of pBR313, using the cycloserine selection procedure developed by Bolivar *et al.* (*11*), followed by in situ hybridization. Table 1 shows the results of such a transformation. No transformants were obtained when *E. coli* cells were transformed with closed circular pBR313 DNA. Neither were transformants obtained when cells were incubated with plasmid DNA converted to linear form by enzymatic cleavage with Bam and digested with S1 nuclease. However, the mixture of plasmid and HbF cDNA which had been treated with T4 ligase gave rise to 60 transformants.

Hemoglobin mRNA from patients with sickle cell anemia was used to synthesize Hb cDNA. This cDNA was then used as a template for the production of ^{32}P-labeled sickle cell cRNA for use as a hybridization probe. Transformed colonies were lysed, fixed onto nitrocellulose filters (*12*), and hybridized with the ^{32}P-labeled cRNA. After hybridization, filters were audioradiographed. Repeated hybridizations showed that several colonies hybridized to the ^{32}P-labeled Hb cRNA (Fig. 3A).

Plasmid DNA was isolated from two colonies that hybridized to [^{32}P]cRNA. These plasmid DNA's were electrophoresed on a 0.7 percent agarose gel with pBR313 DNA (Fig. 3B). On this gel, plasmid DNA's from transformed colonies showed slower mobilities. This indicates plasmids of larger molecular weight, as one would expect after insertion of Hb cDNA into the plasmid.

With the issuance of the National Institutes of Health guidelines, further characterization of these colonies and plasmid DNA's was postponed until the certification of a EK2 host vector system and the official certification of a P3 facility. However, this research was conducted in a limited-access laboratory with negative pressure. All microbiology was done in that laboratory, in a special biocontainment hood meeting the specifications of the NIH guidelines.

During the period when this research was conducted, the laboratory was monitored twice weekly for containment of the host bacteria. At no time was any evidence found for escape of the bacteria into the laboratory environment. Fecal samples were taken throughout the study and found to be free of resistance markers used in this study.

JOHN T. WILSON
Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510
BERNARD G. FORGET, LOIS B. WILSON
Department of Internal Medicine, Yale University School of Medicine
SHERMAN M. WEISSMAN
Department of Human Genetics, Yale University School of Medicine

References and Notes

1. C. A. Marotta, B. G. Forget, M. Cohen-Solal, J. T. Wilson, S. M. Weissman, *J. Biol. Chem.*, in press.
2. M. Cohen-Solal, B. G. Forget, W. Prensky, C. A. Marotta, S. M. Weissman, *ibid.*, in press.
3. C. A. Marotta, J. T. Wilson, B. G. Forget, S. M. Weissman, *ibid.*, in press.
4. For base residue assignments, N signifies any one of the common deoxynucleotide triphosphates, R a purine residue, and Y a pyrimidine. Other abbreviations used are G, guanine; U, uridine; A, adenine; and C, cytosine.
5. B. G. Forget, C. A. Marotta, S. M. Weissman, I. M. Verma, R. P. McCaffery, D. Baltimore, *Ann. N.Y. Acad. Sci.* **241**, 290 (1974); C. A. Marotta, B. G. Forget, S. M. Weissman, J. M. Verma, R. P. McCaffrey, D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2300 (1974); B. DuBuy and S. M. Weissman, *J. Biol. Chem.* **246**, 747 (1971); B. G. Forget and S. M. Weissman, *ibid.* **244**, 3148 (1969); C. A. Marotta, F. Varricchio, I. Smith, S. M. Weissman, M. L. Sogin, N. R. Pace, *ibid.* **251**, 3122 (1976); G. G. Brownlee, *Determination of Sequences in RNA* (Elsevier, New York, 1972).
6. H. F. Bunn, B. G. Forget, H. M. Ranney, *Human Hemoglobin* (Saunders, Philadelphia, 1977).
7. T. Maniatis, S. G. Kee, A. Efstratiadis, F. C. Kafatos, *Cell* **18**, 163 (1976); T. H. Rabbits, *Nature (London)* **260**, 221 (1976); F. Rougeon, P. Kourilsky, B. Mach, *Nucleic Acid Res.* **2**, 2365 (1976); K. O. Wood and J. C. Lee, *ibid.* **3**, 1961 (1976).
8. R. Higuchi, G. U. Paddock, R. Wall, W. Salser, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3146 (1976).
9. B. G. Forget *et al.*, *Cell* **7**, 323 (1976).
10. F. C. Kafatos, A. Efstratiadis, A. M. Maxam, T. Maniatis, *ibid.*, p. 279.
11. F. Bolivar *et al.*, in preparation.
12. M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
13. We thank R. Milardo for technical assistance and J. Allen and B. Holloman for their helpful advice.

7 February 1977

Cloned Ribosomal RNA Genes from Chloroplasts of *Euglena gracilis*

Abstract. *Fragments of* Euglena *chloroplast DNA generated by endonuclease R·Eco RI were separated by agarose-gel electrophoresis into 24 distinct bands. At least five fragments contain sequences complementary to chloroplast ribosomal RNA. Most of the Eco RI fragments have been cloned in a plasmid of* Escherichia coli. *Three of the cloned fragments were shown to contain chloroplast ribosomal RNA sequences by DNA-RNA hybridization.*

Chloroplasts from the unicellular alga *Euglena gracilis* contain multiple copies of a circular, double-stranded DNA molecule (Chl-DNA) whose molecular weight is about 92×10^6 (*1*). These chloroplasts also contain ribosomes with sedimentation values similar to those of prokaryotic ribosomes (*2, 3*), and hybridization of chloroplast ribosomal RNA (Chl-rRNA) to Chl-DNA shows that Chl-DNA contains about two to three rRNA cistrons per chloroplast genome (*3, 4*). The existence of DNA and a protein-synthesizing apparatus in chloroplasts poses some of the most intriguing problems of modern cell biology. How did these organelles arise, and what are their genetic capabilities? How does the nuclear genome interact with the cytoplasmic genome to regulate the biogenesis of the organelle and its energy-producing reactions? Recombinant DNA methodology provides an approach to answer these questions, for it enables us to dissect the chloroplast genome and to determine the genes contained on specific segments.

As a first step toward understanding the organization of chloroplast genes and their functions, we have cloned most of the chloroplast DNA as discrete fragments inserted in a set of RSF2124 plasmids of *Escherichia coli*. These fragments were generated by treatment of Chl-DNA with the restriction endonuclease Eco RI (*5*) and were identified by agarose-gel electrophoresis, which separates DNA molecules according to size (*6*). The purpose of this report is to show that molecular cloning permits unambiguous identification of unique Chl-DNA fragments containing Chl-rRNA genes.

Electrophoresis of Eco RI–digested Chl-DNA on agarose gels yields 24 distinct bands (Fig. 1a). We were able to separate several bands of similar mobility which had not been resolved (*4, 7*). In particular, bands D and E can be separated, and therefore were given individual letter designations. However, bands I, M, and W each contain two DNA fragments so nearly identical in size that we were unable to separate them. These bands are designated I_1, I_2, and so forth. The staining intensity of these bands indicates that each contains more than one