nants. We verified the absence of the λC fragment (see Fig. 2) in all three recombinants by confirming the lack of *int* and *xis* gene function in the red-plaque test of Enquist and Weisberg (*14*). In addition, plating tests showed that all three recombinants failed to grow ($< 2 \times 10^{-10}$) on a *sup*0 host bacterium. These results indicated that the safety features included in the outer arms of the vector phage DNA had been conserved. These data are all consistent with simple insertion of the 5×10^6 dalton chicken rDNA fragment between the purified left and right arms of the vector phage DNA.

We determined the direction of insertion of the rDNA fragment in each recombinant by separating the left and right strands of the hybrid phage DNA by electrophoresis (*15*) and then hybridizing with 18*S* plus 28*S* rRNA (*8*). For λGd1 and λGd3 the rRNA hybridized with the right strand; for λGd2 hybridization was with the left. Thus, transcription of the rRNA sequences should be under control of the early λ phage promoter in λGd2 and of the late phage promoter in λGd1 and λGd3. Figure 2 is a model which indicates the probable structural relationships between the vector phage DNA, recombinant DNA, and chicken rRNA genes.

The recombinants reported here are being used to study the processing of precursor RNA to mature 18*S* and 28*S* rRNA. Cloning of the fragment of molecular weight 12 to 14×10^6 will provide a DNA segment with alternating transcribed and nontranscribed regions, which should serve as a useful model for understanding the organization and regulation of other eukaryotic genes.

Cloning the rRNA gene has provided insight into the requirements for cloning other genes, including those present in a single copy. The fraction of chicken DNA used to construct these recombinants was approximately 5 to 10 percent of the total genome. Three out of about 200 (1.5 percent) of the recombinants contained the 5×10^6 dalton rDNA fragment, which suggests that this fragment makes up about 0.075 to 0.15 percent of the total chicken genome. Knowing the amount of DNA per cell [2.4×10^{-6} μg (*16*)] and the minimal molecular weight of the rDNA unit (17×10^6), we estimate from our data that there are between 215 and 430 copies of the rRNA genes per chicken genome. These values are in line with the data of Sinclair and Brown (*4*), which suggest that there are approximately 175 to 350 copies of the rRNA genes, and with the estimate of Baluda and co-workers (*17*) that there are 210 to 218 copies of the 28*S* gene. Therefore, in our experiments, the production of a specific recombinant was proportional to the fraction of DNA of interest present in the whole genome. This indicates that the cloning of a single-copy gene will require approximately a 200-fold amplification of the methods of detection or gene purification employed in this study.

W. McCLEMENTS
A. M. SKALKA
Roche Institute of Molecular Biology, Nutley, New Jersey 07110

References and Notes

1. J. Morrow, S. Cohen, A. Chang, H. Boyer, H. Goodman, R. Helling, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1743 (1974).
2. R. A. Kramer, J. R. Cameron, R. W. Davis, *Cell* **8**, 227 (1976).
3. L. Enquist, D. Tiemeier, P. Leder, R. Weisberg, N. Sternberg, *Nature (London)* **259**, 596 (1976).
4. J. Sinclair and D. Brown, *Biochemistry* **10**, 2761 (1971).
5. I. Merits, W. Schulze, L. Overby, *Arch. Biochem. Biophys.* **115**, 197 (1966).
6. P. Wensink and D. Brown, *J. Mol. Biol.* **60**, 235 (1971).
7. P. Wellauer and I. Dawid, *ibid.* **89**, 379 (1974).
8. E. Southern, *ibid.* **98**, 503 (1975).
9. R. Helling, H. Goodman, H. Boyer, *J. Virol.* **14**, 1235 (1974).
10. P. Wellauer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2823 (1974).
11. P. Wellauer, I. Dawid, D. Brown, R. Reeder, *J. Mol. Biol.* **105**, 461 (1976).
12. M. Thomas, J. Cameron, R. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4579 (1974).
13. D. Tiemeier, L. Enquist, P. Leder, *Nature (London)* **263**, 526 (1976).
14. L. Enquist and R. Weisberg, *Virology* **72**, 147 (1976).
15. G. Hayward, *ibid.* **49**, 342 (1972).
16. M. Baluda, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 576 (1972).
17. M. Shoyab, P. Markham, M. Baluda, *J. Virol.* **14**, 225 (1974).
18. M. Gross-Bellard, P. Dudet, P. Chambon, *Eur. J. Biochem.* **36**, 32 (1973).
19. R. Laskey and A. Mills, *ibid.* **56**, 335 (1975).
20. N. Arnheim and E. M. Southern, *Cell*, in press.
21. I. Dawid and P. Wellauer, *ibid.* **8**, 443 (1976).
22. D. Gillespie and S. Spiegelman, *J. Mol. Biol.* **12**, 829 (1965).
23. J. R. Cameron, S. M. Panasenko, I. R. Lehman, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3416 (1975).
24. The T4 ligase was a gift of C. Harvey, Hoffmann-La Roche.
25. National Institutes of Health guidelines for recombinant DNA research, *Fed. Regist.* **41** (No. 131), 27917 (1976). This project was approved by an Institutional Biohazards Committee constituted and functioning as stipulated in these NIH guidelines.

3 January 1977

Clones of Individual Repetitive Sequences from Sea Urchin DNA Constructed with Synthetic Eco RI Sites

Abstract. *Interspersed repetitive sequences were isolated from sea urchin DNA by renaturing to low* C_0t *followed by treatment with nuclease S1. Synthetic Eco RI sites were ligated onto the repetitive sequence elements, which were then inserted at the Eco RI site of plasmid RSF2124 and cloned. The repetitive sequences can be excised from the plasmid with Eco RI for further study.*

To facilitate studies of individual repetitive DNA sequence families, we have constructed recombinant DNA plasmids containing repetitive DNA from *Strongylocentrotus purpuratus* (*1*). Our object was to obtain cloned sequences terminated at the ends of interspersed repetitive sequence elements (*2*), rather than where restriction enzyme sites happen to fall. This was accomplished through the use of synthetically prepared Eco RI restriction sites that were ligated to the repetitive DNA fragments. The product of the ligation reactions was cloned to provide a source of plasmids from which individual repetitive DNA sequences could be reisolated in relatively large amounts.

Repetitive DNA duplex was prepared as follows (details may be found in the legend to Fig. 1). Sea urchin DNA sheared to a single-strand weight mean length of about 2000 nucleotides was renatured to C_0t 40 (C_0 is the initial concentration of nucleotides in moles per liter, and t is the time in seconds) and then treated with single-strand specific S1 nuclease. Previous studies (*2, 3*) have shown that by this point most repetitive sequences in the sea urchin genome are fully renatured. The nuclease digestion removes the nonrepetitive regions flanking the interspersed repetitive sequences, since these remain single-stranded at C_0t 40 (*2*). The resistant DNA sequence consists almost entirely of repetitive duplexes. We have shown earlier that a majority of the S1 nuclease resistant duplexes are 300 to 400 nucleotides in length, and that this is the characteristic length of interspersed repeats in the sea urchin genome (*2, 3*). The resistant repetitive duplexes are expected to have 3′-hydroxyl and 5′-phosphoryl termini (*4*).

A symmetrical decamer containing the Eco RI restriction site was synthesized by a triester chemical synthesis method (*5*). The structure of this decamer is shown in Fig. 1. The 5′-hydroxyl termini of the Eco RI decamers were labeled with ^{32}P, using T4 polynucleotide kinase. The decamer was covalently linked to the repetitive sea urchin DNA duplexes by blunt-end ligation with T4 DNA ligase (*6*). The expected products of the blunt-end ligation are the repetitive sea urchin DNA fragments bearing covalently linked Eco RI decamers, plus polymerized Eco RI decamers. The ligated mixture was next digested with Eco RI endonuclease. Eco RI treatment should yield the repeti-

tive sea urchin DNA with covalently linked Eco RI sticky ends labeled with ^{32}P, plus Eco RI cleavage products of polymerized and monomeric decamer. Sea urchin DNA containing covalently attached ^{32}P-labeled Eco RI cohesive ends was separated from the other cleavage products by gel filtration. From the amount of ^{32}P associated with the sea urchin DNA in the exclusion peak, the specific activity of the Eco RI decamer, and the DNA mass recovered, we calculated that 30 percent of the ends of the sea urchin DNA molecules were covalently linked to an Eco RI site. Thus about 10 percent of the repetitive DNA duplexes could be expected to carry Eco RI sites on both ends. This preparation was reacted with Eco RI–cleaved DNA isolated from the plasmid RSF2124, with T4 DNA ligase included to seal the duplexes formed from the Eco RI cohesive ends. The procedures used to prepare the repetitive DNA and to construct the clones is illustrated in Fig. 1.

RSF2124 is a colicinogenic (E1) plasmid carrying a locus for ampicillin resistance (*amp*^{R}) and a single Eco RI site (*7*). Transfection of *Escherichia coli* strain C600 with the recombinant plasmids was accomplished essentially according to Cohen *et al.* (*8*). Transformation efficiency was from 10^{-7} to 10^{-6} Amp^R colonies per molecule of plasmid DNA. The single Eco RI site in the RSF2124 genome lies within the colicin structural gene or in a control region for colicin E1 production (*7*). Insertion of a segment of DNA into the plasmid genome with the use of the Eco RI cohesive ends thus results in dysfunction of the colicin E1 locus. This property was exploited to assay the Amp^R colonies for insertion of the sea urchin DNA fragments into the plasmid genome. A double-layer technique (*9*) was used to detect colicin-producing clones, as described in the legend to Fig. 1. The frequency of Amp^R clones which were *col*$^-$ was about 10 percent.

Insertion of the Eco RI–sea urchin DNA fragments in the RSF2124 genome should introduce a second Eco RI site into the hybrid plasmid genome. Eco RI cleavage of hybrid plasmid DNA should thus result in unit length linear plasmid DNA molecules plus repetitive DNA fragments. We examined a number of *amp*^{R}*col*$^-$ clones and found that more than 65 percent contain short DNA duplexes which can be excised from the plasmid genome with Eco RI. We consider this a minimum estimate because of occasional difficulties encountered in detecting small inserted fragments representing 3 percent or less of the DNA mass. In other experiments in which long-

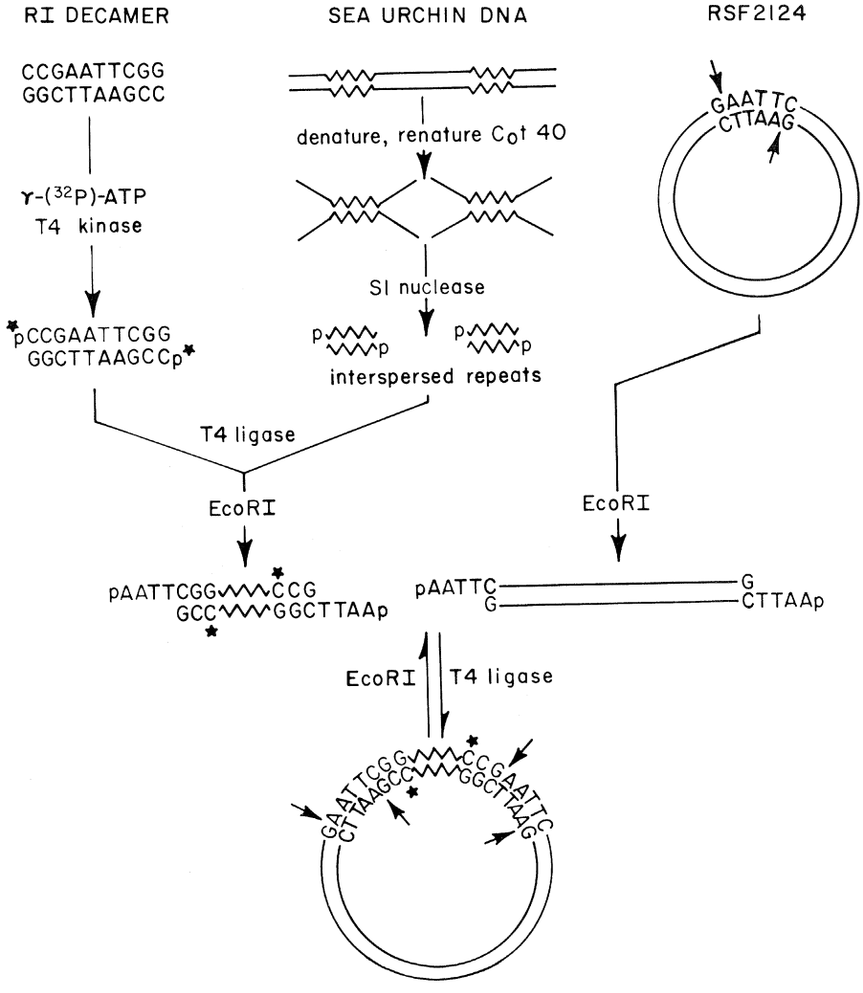

Fig. 1. Enzymatic steps in the construction of recombinant plasmids containing repetitive sea urchin DNA. Repetitive sea urchin DNA (double wavy lines), Eco RI decamer (shown in its entirety), and RSF2124 DNA (double straight lines) were ligated and cloned as follows. The vertical arrows (⇅) indicate sites of Eco RI cleavage. The star (★) indicates the ^{32}P label at the 5′ termini. Repetitive sea urchin DNA was isolated by renaturing 2000-nucleotide long sperm DNA to C_0t 40, followed by digestion of the single-strand tails with 200 μl of S1 nuclease per milligram of DNA at 37°C for 45 minutes (*3*). Ten nanomoles of Eco RI decamer were 5′ phosphorylated with 7 units of T4 polynucleotide kinase and 50 nanomoles of [γ-^{32}P]ATP (adenosine triphosphate) for 1 hour at 37°C in 0.05*M* tris-HCl (*p*H 9.0), 0.01*M* $MgCl_2$, 0.005*M* dithiothreitol (DTT). A 60-fold excess of ^{32}P-labeled Eco RI decamer (16 μg) was ligated for 17 hours at room temperature to the repetitive DNA fragments (8 μg) with 10 μl of T4 DNA ligase in 0.04 ml, 66 m*M* tris-HCl (*p*H 7.6), 6.6 m*M* $MgCl_2$, 10 m*M* DTT, 0.01 percent NP40, and 0.4 m*M* ATP. This reaction mixture was adjusted with 100 m*M* tris-HCl (*p*H 7.6), 50 m*M* NaCl, 10 m*M* $MgCl_2$, and 0.02 percent NP40 and digested with 15 μl Eco RI endonuclease for 3 hours at 37°C. The Eco RI digest was deproteinized with an equal volume of $CHCl_3$ and isoamyl alcohol (24 : 1). The cleavage products were resolved on a Sephadex G-200 column. Supercoil RSF2124 plasmid DNA, cleaved with Eco RI essentially as described above, was mixed with a 40-fold molar excess of G-200–excluded sea urchin DNA in 66 m*M* tris-HCl (*p*H 7.6), 20 m*M* NaCl, 1 m*M* EDTA, 10 m*M* $MgCl_2$, 10 m*M* DTT, and 66 μ*M* ATP in a total volume of 30 μl. One unit of T4 DNA ligase (Miles Laboratories) was added, and the reaction mixtures were incubated at 15°C for 18 hours. Five microliters of this reaction mixture, diluted 200-fold with 0.05*M* $CaCl_2$, was used to transform *E. coli* strain C600 (Rec$A^+r^-m^-$) to ampicillin resistance essentially according to Cohen *et al.* (*8*). Transformed cells were plated on agar plates containing ampicillin (15 μg/ml). Amp^R colonies were assayed for colicin production by a double-overlay method (*9*). Colonies were transferred to duplicate plates and grown overnight. One plate was retained as a master, and the cells on the second plate were killed with $CHCl_3$. The latter plates were overlaid with top agar containing 1×10^8 colicin-sensitive bacteria (*E. coli* HB101). The plates were incubated overnight and then scored for the absence of colicin production (*col*$^-$). Col$^+$ colonies are surrounded by a clear halo, indicating lack of bacterial growth, while col$^-$ colonies are overgrown by the colicin-sensitive cells. Col$^-$ colonies were stored in stab cultures and assayed for the presence of an inserted sea urchin fragment.

er sea urchin DNA fragments derived from an Eco RI digest of sea urchin DNA were inserted into the same plasmid vector, about 80 percent of the $amp^{R}col^{-}$ clones were found to contain sea urchin DNA.

The lengths of the inserted repetitive DNA fragments were estimated by gel electrophoresis, with the use of an Hae III digest of the parent plasmid as a size standard. The lengths of the inserted DNA sequences from 20 clones have now been measured. Of these, 17 are less than 1000 nucleotides. The mode value of their length distribution is 300 to 400 base pairs. Two inserts are in the range of 1000 to 1200 nucleotides, and one is about 2000 nucleotides. Although the sample is small, this distribution is in excellent agreement with the distribution of fragment lengths in the original nuclease S1 digest and with the range of repetitive sequence lengths in the sea urchin genome. More than 90 percent of the repetitive sequence elements in this genome are 300 to 400 nucleotides long, with less than a few percent $\geq$ 2000 nucleotides. Our data thus suggest that the population of repetitive sea urchin DNA duplexes inserted into the RSF2124 vector is reasonably representative of the repetitive DNA sequences in the sea urchin genome.

To verify that the DNA sequences inserted into the plasmid genome in fact consist of repetitive sea urchin DNA, the cloned inserts were reassociated with sheared sea urchin DNA. The following procedure was used. The cloned DNA fragments were excised with Eco RI and labeled at the 5′ termini with ^{32}P by the T4 kinase method (*10*) (Fig. 2). The DNA was renatured after labeling and placed on 3 percent agarose gels along with an Hae III digest of RSF2124. An example is shown in Fig. 2A, which illustrates the preparation of the repetitive DNA insert from clone CS2108. After renaturation, the labeled DNA insert from this clone migrates to a position corresponding to a 190 base pair fragment, where it can be detected either by ethidium bromide staining or by autoradiography. The labeled DNA was eluted from the gel and reassociated with excess sheared sea urchin DNA approximately 500 nucleotides in length. The fraction of DNA reassociated was determined by hydroxyapatite chromatography and is plotted as a function of the driver DNA C_0t in Fig. 2B. The self-reaction of the labeled CS2108 fragment was monitored in separate reactions (not shown). At the tracer concentration used, the rate of the self-reaction was less than 10 percent of the driver reaction and thus does not affect the results. A least-squares solution to the kinetic data reproduced in Fig. 2B yields a rate constant of $3.16 \times 10^{-2} M^{-1}$ sec^{-1} for the reassociation of the CS2108 fragment with sea urchin DNA. The rate constant for the reassociation of single-copy sea urchin DNA of similar fragment length with whole DNA is $1.13 \times 10^{-2} M^{-1}$ sec^{-1}. Thus clone CS2108 contains a DNA sequence which is repeated approximately 25 times per haploid genome. It is unlikely that the reiteration frequency of the CS2108 sequence could differ from this number by more than a factor of two or three, taking into account the limited kinetic data and uncertainties in correcting the observed rate constants for minor length differences and the effect of base pair mismatch.

The repetitive sea urchin DNA clones constructed by the blunt-end ligation

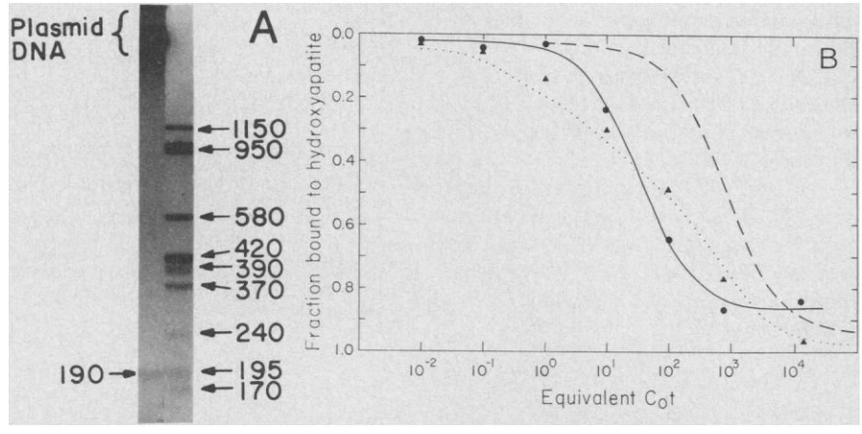

Fig. 2. Analysis of a recombinant plasmid clone constructed by blunt-end ligation. (A) Detection and end labeling of inserted sea urchin fragment. $Amp^{R}col^{-}$ clone CS2108 was grown in 40 ml of L broth plus ampicillin (15 μg/ml) to 4×10^{8} to 6×10^{8} cells per milliliter, and the plasmid content was amplified by incubation overnight in chloramphenicol (200 μg/ml). The cells were harvested by centrifugation and washed with a solution of 10 m*M* tris-HCl (*p*H 7.4) and 1 m*M* EDTA. They were again centrifuged and resuspended in 3 ml of 25 percent sucrose, 50 m*M* tris-HCl (*p*H 8.0), and 40 m*M* EDTA at 0°C. Lysozyme (5 mg; Sigma) was added, and the solution was maintained at 0°C for 5 minutes. Then 0.5 ml of 0.5*M* EDTA and 0.375*M* EGTA (*p*H 8.0) was added for 5 minutes at 0°C. Four ml of Brij lysing buffer [1 percent Brij 58, 0.4 percent sodium deoxycholate, 0.063*M* EDTA, 0.05*M* tris-HCl (*p*H 8.0)] was added and mixed vigorously. Chromosomal DNA was removed from the lysate by centrifugation at 50,000 rev/min for 45 minutes (Beckman/Spinco 50 Ti rotor). Superhelical plasmid DNA was isolated from the cleared supernatant by CsCl–ethidium bromide density gradient centrifugation (ethidium bromide at 250 μg/ml; ρ_0 = 1.669 g/cm^3) in a 50 Ti rotor at 39,000 rev/min. Superhelical DNA was removed from the gradient by side puncture. Ethidium bromide was removed by extraction with isopropanol. Plasmid DNA was dialyzed against 100 m*M* tris-HCl (*p*H 7.5), 100 m*M* NaCl, 5 m*M* $MgCl_2$ at 4°C. Recombinant plasmid DNA (15 μg) was digested with 2 μl of Eco RI at 37°C for 30 minutes. This digest was concentrated by flash evaporation to 0.5 ml and passed over a 0.5-ml Chelex 100 (BioRad) column. The DNA was precipitated at −70°C for 2 hours with 1 volume of isopropanol. The precipitate was collected by centrifugation at 10,000*g* for 30 minutes. The dried precipitate was resuspended in 10 m*M* tris-HCl (*p*H 8.0) and reacted with 1 μl of bacterial alkaline phosphatase (Worthington) at 37°C for 30 minutes. Ten micrograms of superhelical plasmid DNA carrier and 1 volume of 0.3*M* sodium acetate (*p*H 6.8) was added. This solution was extracted twice with a mixture of 80 percent phenol (*p*H 8.0) and 20 percent chloroform: isoamyl alcohol (24 : 1), and once with ether, then precipitated as described above. Labeling at 5′ termini was done according to Maxam and Gilbert (*10*) with [γ-^{32}P]ATP, specific activity 1200 c/mmole (Winston Salser's laboratory, University of California, Los Angeles). Kinase reactions were deproteinized as described above, and precipitated with 2.5 volumes of 100 percent ethanol at −70°C as described above. The dried precipitate was resuspended in a volume of 200 μl of 0.1*M* NaCl, 0.05*M* tris-HCl (*p*H 7.2). The concentration of the end-labeled CS2108 DNA fragment was determined by its rate of renaturation, and from this the specific activity was calculated to be about 1×10^{6} to 2×10^{6} count/min per microgram. A portion of the labeled DNA was heat denatured and renatured at 60°C overnight. The renatured DNA fragment was loaded on a 3 percent agarose gel and subjected to electrophoresis (gel 1) along with an Hae III digest of RSF2124 DNA (gel 2) at 60 volts for 3.5 hours. Gels were stained with ethidium bromide (1 μg/ml). Lengths of the Hae III fragments of RSF2124 are indicated in the figure. (B) Reassociation of end-labeled Eco RI fragment from a blunt-end repeat clone with excess sea urchin DNA. The 190 base pair end-labeled CS2108 fragment from (A) was eluted from the gel by dissolving in 5*M* $NaClO_4$ and binding the DNA to hydroxyapatite in 0.12*M* phosphate buffer (PB), followed by thermal elution of the ^{32}P-labeled DNA. This DNA was reassociated with excess 500 nucleotide long (5×10^{5} mass ratio) sheared sea urchin DNA in 0.12*M* PB at 55°C. The fraction ^{32}P-labeled DNA (●) and driver DNA (▲) reassociated was determined by hydroxyapatite chromatography in 0.12*M* PB at 55°C. The solid line is the least-squares solution for the reaction of ^{32}P-labeled DNA driven by total sea urchin DNA. The rate constant determined for this reaction is $3.16 \times 10^{-2} M^{-1}$ sec^{-1}. The dashed line illustrates the reassociation of nonrepetitive sea urchin DNA with whole DNA (data not shown). The rate constant for this reaction is $1.13 \times 10^{-3} M^{-1}$ sec^{-1}. The dotted line shows the renaturation of total sea urchin DNA (*2*).

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.