Volume 11 Number 16 1983 Nucleic Acids Research

# An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid

Li-He Guo, Robert C.A.Yang+ and Ray Wu*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 25 April 1983; Revised and Accepted 19 July 1983

ABSTRACT

A strategy for kilo-base sequencing of a target DNA cloned in plasmid pWR34 is described. A long target DNA is progressively shortened from one end, by digestion with BAL31 nuclease or exonuclease III and nuclease S1, followed by cleaving off the shortened vector DNA. The family of the shortened target DNA molecule is next cloned in between the StuI site on one end, and a cohesive-ended restriction site on the other end, within the polylinker region of pWR34. DNA fragments cloned into this plasmid are sequenced directly by using a synthetic oligonucleotide primer, which binds to one side of the polylinker region using the dideoxynucleotide chain-termination method. The plasmid DNA, easily obtained by adoption of a rapid mini-preparation, is usually pure enough for direct DNA sequencing. Thus, both strands of any DNA several thousand base pairs in length can be completely sequenced (using two different primers) with ease within a short time, without the need for constructing a physical map.

## INTRODUCTION

Knowledge of the nucleotide sequence of DNA is essential in understanding gene organization and expression at the molecular level. The chain-termination method of Sanger et al[1], and the chemical modification method of Maxam and Gilbert[2] have been extensively adopted for sequencing DNA. In order to increase the ease and speed of cloning and sequencing genes, several improved methods based on the use of single-stranded phage M13[3-7] or double-stranded plasmids[8-11] have been developed. Rapid sequencing of short DNA fragments (shorter than 500 bp) by the chemical method using pUR222[8] or pUR250[29], and by the chain-termination method using pWR2[11] or pWR32[16] have been described.

Several strategies for rapid sequencing of long DNA fragments, which are over 2000 base pairs in length, have become available recently using either the phage M13 system[12-15] or the plasmid system[16,17].

In this paper, we describe an improved vector and a simple strategy for cloning the progressively shortened DNA into the cloning vector.

Both strands of the cloned DNA can then be sequenced directly and completely. Using this method, we report the result of sequencing a 2.2 kb long human DNA fragment, 1 kb of which was obtained on a single sequencing gel.

## MATERIALS AND METHODS

### Materials

*E. coli* exonuclease III, nuclease S1 and BAL31 nuclease were from Bethesda Research Laboratories, Inc. *E. coli* DNA polymerase (large fragment) and restriction enzymes were from New England Biolabs. Recently, a more highly purified large fragment of *E. coli* DNA polymerase has become available from Bethesda Research Laboratory (cat. No. 8012 SA/SB). This fragment comes from cells carrying a cloned gene which codes for only the large fragment of DNA polymerase. T4 DNA ligase was from Boehringer Mannheim.

The [$\alpha$-$^{32}$P]dNTPs (410 Ci/mmol, 1 mCi/ml) were from Amersham Corporation. The dNTPs and ddNTPs were from P-L Biochemicals, Inc. Synthetic *Stu*I adaptor d(C-A-G-G-C-C-T-G) was from Worthington Diagnostic Systems. Primer I, d(A-C-C-A-T-G-A-T-T-A-C-G-A-A-T-T)[11,17], was synthesized by F. Georges and S.A. Narang before it was commercially available. A primer similar to the primer I can now be obtained from New England Biolabs. Primer II, d(T-C-C-C-A-G-T-C-A-C-G-A-C-G-T), was purchased from New England Biolabs.

Agarose and low-melting-point agarose were from Bethesda Research Laboratories, Inc.

Plasmid p27 carrying a 2.2 kb long human DNA cloned in the *Eco*RI of pBR322 was supplied by Dr. Ru-Chih C. Huang and Alan Deutch. Plasmid pWR34 was constructed by insertion of a *Stu*I linker into the *Bam*HI site in pWR32[16].

Cells from *E. coli* JM101, JM83, JM83($r^-$) or RR1$\Delta$M15 were made competent and stored frozen, according to Morrison[17].

Buffers or media used are as follows: exonuclease III buffer: 66 mM Tris-HCl, pH 8, 77 mM NaCl, 5 mM $MgCl_2$, 10 mM dithiothreitol (DTT). 10 x S1 buffer: 0.5 M NaOAc, pH 4, 0.5 M NaCl, 60 mM $ZnSO_4$. BAL31 buffer: 20 mM Tris-HCl, pH 8, 12 mM $MgCl_2$, 12 mM $CaCl_2$, 600 mM NaCl, 1 mM EDTA. TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8. T4 ligase buffer: 50 mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$, 5 mM DTT, 0.5 mM ATP. RE buffer: 50 mM Tris-HCl, pH 7.6, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT. 2 x YT medium:

16 g of tryptone, 10 g of yeast extract, 5 g of NaCl in 1000 ml of deionized or glass-distilled water. YT agar medium: 15 g bactoagar, 8 g of tryptone, 5 g of yeast extract, 5 g of NaCl in 1000 ml of deionized or glass distilled water.

Methods

(A) Methods for the digestion of DNA with a restriction enzyme, electrophoresis on a low-melting-point agarose gel, ligation, transformation and preparation of plasmid DNA and its derivatives have been described in detail[11].

(B) Digestion of DNA with BAL31 nuclease, or with E. coli exonuclease III followed by nuclease S1 was carried out according to our previous papers[10,11]. In some experiments, BAL31 nuclease digestion of DNA was carried out at 37° instead of 23°, and the ratio of enzyme to DNA concentration was changed from 1:1 to 0.25:1; under these conditions, the rate of degradation of DNA from each end of DNA was 100-150 nucleotides per minute.

For progressive shortening of the target DNA with E. coli exonuclease III and nuclease S1, 36 µg of the p27-34 DNA was first linearized with HindIII (Figure 1) in 120 µl of RE buffer, for subsequent sequencing of the upper strand of the 2.2 kb long target DNA. After completion of HindIII digestion, 48 µl of 10 x exonuclease III buffer[10], 312 µl of water and 660 units of exonuclease III were added, and incubation was carried out at 23°. Aliquots were pipetted out at different times: 40 µl at 20 min, 80 µl at 40 min, 160 µl at 60 min, and 200 µl at 80 min. To each aliquot, 10 x S1 buffer[20] was added immediately (4.5 µl, 9 µl, 18 µl, or 22.5 µl, respectively) to terminate the reaction. Nuclease S1 was added (60, 120, 240, or 300 units, respectively), and incubation was carried out at 23°C for 15 min. A stop solution of 0.1 M EDTA-1.5 M NaOAc, equal to 1/5 volume of each aliquot, was then introduced to each reaction mixture. If necessary, a portion of the DNA from each mixture was ethanol precipitated and electrophoresed on an 1% agarose to analyze the progress of digestion.

For progressive shortening of the HindIII linearized target DNA with BAL31 nuclease, 36 µg of HindIII-treated p27-34 DNA in 120 µl of RE buffer was digested with 36 units of BAL31 nuclease at 23°C, after addition of 96 µl 5 x BAL31 buffer[11] and 240 µl $H_2O$. Aliquots of 40 µl, 80 µl, 160 µl and 200 µl were separately removed at 2 min, 4 min, 6 min and 8 min. The reaction was terminated by introduction of 1/20 volume

of 0.5 M EDTA (pH 8). For analyzing the progress of digestion, a portion of the DNA from each aliquot was precipitated with 2.5 volume of ethanol (chilled at -70° for 5 min, centrifuged for 5 min and dried for 5 min in a vacuum), and resuspended in 8 µl of TE buffer. The DNA was electrophoresed on a 1% agarose gel, using EcoRI and HindIII-treated λ DNA as size markers.

(C) In preparation for subcloning and sequencing the target DNA, the samples after exonuclease III and nuclease S1 digestion or after BAL31 nuclease digestion, were extracted once with an equal volume of phenol saturated with 1 M Tris-HCl, pH 8, and chloroform-isoamyl alcohol (24:1, v/v), respectively. After ethanol precipitation, washing and drying, the DNA pellet was resuspended in 20 µl of RE buffer[11], followed by EcoRI and PvuII digestion if the plasmid p27-34 was linearized with HindIII (or by HindIII and SmaI digestion if the plasmid DNA was linearized with PvuII). Each sample was electrophoresed on a 1% low-melting-point agarose gel and the gel fragments carrying the shortened target DNA of desired lengths were excised.

(D) Subcloning the shortened target DNA fragments into plasmid pWR34 was carried out as follows. The vector pWR34 was digested (step e) first with StuI, and the reaction mixture was heated at 70°C for 10 min to inactivate StuI and the contaminating enzyme. After cooling the mixture to room temperature, EcoRI (or HindIII if the plasmid DNA was linearized with PvuII) was added and incubation was continued at 37°C. The sample was electrophoresed (step d), together with the shortened target DNA samples, on the same 1% low-melting-point agarose to separate the large fragment of the vector from the small. Each desired DNA fragment was cut out of the gel, incubated at 70° for 2 min to melt the gel, diluted with 1/10 volumes of 5 M NaCl, heated for an additional 3 min, extracted twice with phenol and once with chloroform. The aqueous layer containing the target DNA was mixed with that of the vector DNA (0.5 µg), and the mixture was precipitated by ethanol, washed once with ethanol and dried. The DNA pellet was resuspended in 10 µl of T4 ligase buffer[11] and 1 unit of T4 DNA ligase (Weiss unit). Incubation was carried out at 4° overnight. To each ligation mixture was added 200 µl of E. coli JM83, JM83($r^-$) or RRIΔM15 competent cells. After incubation at 0° for 30 min, and then at 42° for 2 min, 1 ml of 2 x YT medium (no ampicillin) was added to each mixture and incubated at 37° for 1 hour. The samples were plated on petri dishes[11] which contained

YT agar medium plus 100 µg/ml ampicillin, and 40 µg/ml each of IPTG and X-gal. The plates were incubated at 37° overnight and those bacteria which gave white colonies were selected.

(E) Mini-preparation of plasmid DNA was carried out according to our previous method[11]. Usually 20-50 µg of plasmid DNA are obtained from bacteria growing on 1/8 of a petri dish containing YT medium plus 100 µg/ml of ampicillin. The DNA sample is usually pure enough for sequencing. Between 4-6 colonies are picked from each suitable time point of the progressively shortened and cloned DNA. The size of the shortened target DNA is determined after digestion with two restriction enzymes, and electrophoresed with appropriate size markers on an 1% agarose gel. For plasmid preparation in liquid medium, M9 medium gives higher yields of plasmids than YT medium[11].

(F) DNA sequence analysis was carried out using the primer method which was described in our earlier paper[11]. In brief, 0.4 pmol of plasmid DNA was linearized with 3-6 units of SmaI or PvuII. After 30-60 min at 37°, 10 pmoles of primer I (A-C-C-A-T-G-A-T-T-A-C-G-A-A-T-T) or primer II (T-C-C-C-A-G-T-C-A-C-G-A-C-G-T) were added, and the mixture was boiled in a sealed capillary for 3 min and immediately cooled to 0°C. The sample was divided into 4 tubes and each incubated with 4 dNTPs (including one with $^{32}P$), and a different ddNTP in the presence of DNA polymerase. One µl of sample was loaded in each slot of the sequencing gel.

## RESULTS

### Construction of pWR34

A DNA of several kilo bases (kb) in length can be sequenced after it is progressively shortened by digestion with exonuclease III followed by S1 nuclease[11,16] or by BAL 31 nuclease[11,14] (Figure 1, step b). The shortened DNA fragments require blunt-end ligation to the HincII-cut vector DNA for subcloning and sequencing. However, the commercial HincII is usually contaminated with an exonuclease, so that white colonies without an inserted DNA can be produced by ligation of the vector DNA after one or several nucleotides are removed by the contaminating nuclease[11,16]. In order to overcome this problem, we have inserted a StuI site into the polylinker region of pWR32, to provide another site for blunt-end ligation of target DNA (Figure 1, step e) to be cloned and sequenced.

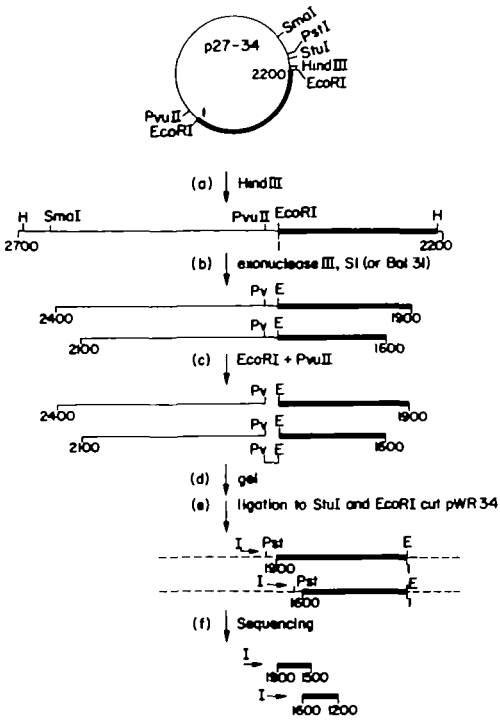



Fig. 1. Diagram illustrating the strategy for kilo-base sequencing of DNA. p27-34 is used as an example to describe the strategy. p27-34 (4.9 kb) contains a 2.2 kb human DNA fragment (dark line) cloned in the EcoRI site of the polylinker region of pWR34. Details for each step are given in the "Result" section. Note that between steps (c) and (f), the orientation of the target DNA is reversed for convenience. Abbreviations for restriction enzyme sites are: E for EcoRI, Pv for PvuII and H for HindIII.

The heavy line represents the target human DNA fragment, the thin line represents the vector DNA, and the dashed line represents the plasmid pWR34 DNA used for subcloning in step (e).

Alternatively, in step (c), PvuII can be omitted if the size of the insert is at least 700 bp shorter than the 2700 bp vector. In this case, agarose gel electrophoresis can distinguish the target DNA fragment from the vector DNA, and only the target DNA is eluted from the low-melting agarose gel for ligation in step (e).

The plasmid pWR34 was constructed by digestion of pWR32 with BamHI, and the cohesive end filled in by repair synthesis[18]. A synthetic StuI adaptor: d(pC-A-G-G-C-C-T-G) was blunt-end ligated to the BamHI-cut and filled-in pWR32 to produce pWR34, which is identical to pWR32 except that 12 base pairs are added to introduce a new StuI site. Since 12 bp were added, the reading frame was maintained for the synthesis of the

α-peptide of β-galactosidase downstream from the lac promoter and the polylinker region. E. coli JM83, JM83($r^-$), JM101, or RR1ΔM15 harboring pWR34 gives blue colonies on indicator plates containing IPTG and X-gal. If a target DNA is inserted into the polylinker region between the lac promoter and the lac Z' gene, bacteria harboring the hybrid plasmid generally gives rise to white colonies. This makes the selection easy for bacteria colonies carrying inserted target DNA[3,8,11]. The physical map of pWR34 is shown on the upper left-hand corner of Figure 2, and the DNA sequence of the polylinker region is shown in Figure 3. There are a total of nine unique restriction enzyme sites in the polylinker region. Only two of them (StuI and HincII) give blunt-ended DNA after cleavage of the vector with these enzymes. For efficiently cloning a target DNA into this vector, it is preferable to cut the vector to produce cohesive ends (e.g. with EcoRI), or to use two enzymes (e.g. StuI and EcoRI) to produce one blunt end and one cohesive end. If using two enzymes to digest a DNA, a blunt-end-producing enzyme should be used first, because the enzyme is often contaminated with an exonuclease. After the enzyme (e.g. StuI) digestion, the reaction mixture is heated at 70° for 10 min., followed by the second enzyme (e.g. EcoRI) digestion.

Construction of p27-34

Before sequence analysis of the 2.2 kb human DNA, plasmid p27 was cleaved with EcoRI to excise the human DNA. After gel fractionation, the 2.2 kb human DNA was ligated to the EcoRI-cut pWR34 to produce p27-34 (Fig. 2). The sequence of the human DNA insert (heavy line) can be analyzed directly in p27-34 by using primer I and primer II. The sequence of approximately 400 bp can be obtained from each primer binding site. However, for analyzing the entire 2.2 kb sequence, another strategy is needed.

Strategy for subcloning and sequencing the 2.2 kb human DNA

Our strategy for sequencing a long DNA such as the 2.2 kb human DNA in plasmid p27-34 includes the following steps, as shown in Figure 1.

(a) The plasmid p27-34 DNA is linearized by digestion with HindIII to expose one end of the target DNA.

(b) The linearized DNA is progressively shortened from both ends with exonuclease III and nuclease S1 (or by BAL31 nuclease) to give blunt-ended DNA. However, only one end of the target DNA (heavy line) is shortened (from the right-hand end). Samples are pipetted out at different intervals, so that approximately 300 bp are removed from each

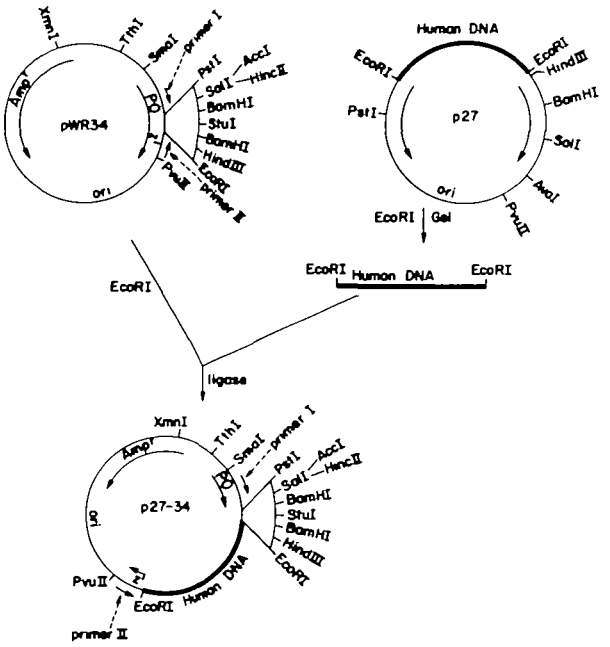



Fig. 2. Construction of plasmid p27-34 carrying a 2.2 kb human DNA fragment. The human DNA fragment inserted in EcoRI site of p27 was excised by digestion with EcoRI, followed by electrophoresis on an 1% low-melting-point agarose gel. After elution from the gel, the 2.2 kb human DNA was cloned into the EcoRI site of pWR34 (2.7 kb). After transformation of E. coli cell JM101, white colonies were selected. DNA from these colonies can be sequenced directly by using synthetic primer I or primer II. They are universal primers because they are located outside the polylinker region in pWR34 that flanks any cloned DNA fragment.

end between time intervals.

(c) Each sample that contains the shortened DNA is digested with EcoRI and PvuII, to cut the DNA into three fragments: the desired target DNA (300-1900 bp) contains one blunt-end and one cohesive end (EcoRI), the large vector DNA contains two blunt ends, and a small fragment of the vector DNA (90 bp) contains one blunt end and one cohesive end.

(d) The restriction enzyme digested samples are fractionated by electrophoresis on 1% low-melting-point agarose gel to purify and select the desired shortened target DNA. The target DNA of sizes corresponding to 1900 bp, 1600 bp, 1300 bp, etc., are cut out from the gel, phenol and chloroform extracted and ethanol precipitated.

(e) The target DNA from step (d) is ligated to StuI- and EcoRI-cut

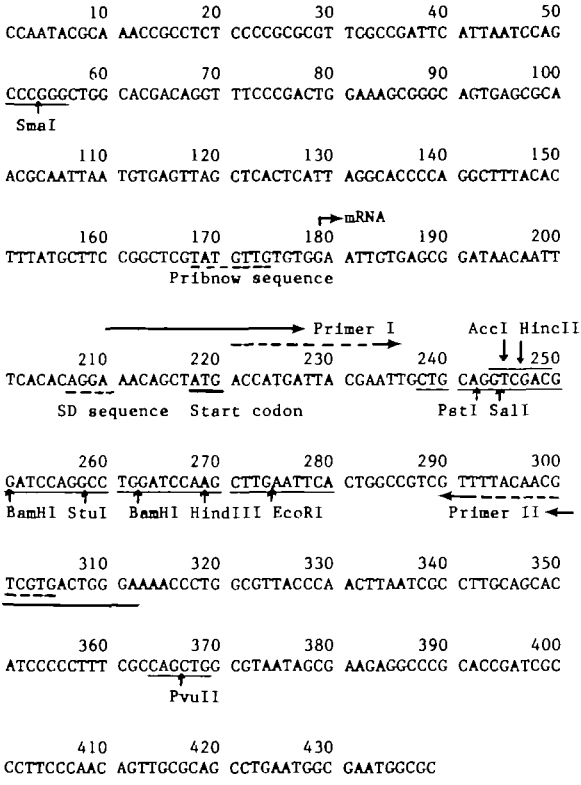



Fig. 3. Nucleotide sequence of lac region in pWR34. Two dotted arrows represent two chemically synthesized primers, and the direction of sequencing DNA. Two solid arrows represent two primers purchased from New England Biolabs (originally designed for the M13 sequencing system). Each primer can be used to determine the sequence of a specific strand of a DNA fragment ligated in between the polylinker region.

vector DNA pWR34. The EcoRI end of the target DNA is ligated with the EcoRI cohesive end of the vector DNA in one orientation only, and the blunt end of the target DNA is ligated to the StuI end of the vector. The original vector DNA from p27-34 now contains two blunt ends, and is thus excluded from ligating to the new vector pWR34. After transformation, the desired clones can be easily selected as white colonies. The StuI and EcoRI treated vector DNA is unable to ligate by itself (except for dimerization), so that the background is low.

Hybrid plasmid DNA were made using a rapid mini-preparation, and four to six clones that represent each specific size class of shortened target DNA were analyzed. The length of the target DNA is confirmed

after digesting a portion of each hybrid plasmid with EcoRI and BamHI to liberate cloned target DNA, followed by gel electrophoresis on an 1% agarose gel. Those clones that carry the desired sizes of target DNA are chosen for sequence analysis.

(f) The hybrid plasmid DNA of approximate size classes is sequenced directly by linearizing the plasmid with SmaI, hybridizing primer I, and extending the primer using the dideoxynucleotide chain termination method. For example, the original clone p27-34 can be sequenced using primer I to give the sequence of nucleotides 2200 to 1800. Subclone I gives nucleotides 1900 to 1500, etc., until the entire upper strand of the target DNA is determined. In this example, the sequence of one clone overlaps approximately 100 nucleotides of the next clone.

The lower strand of the target DNA can be sequenced in the same way except that PvuII is used in step (a), HindIII and SmaI (or BamHI) are used in step (c), StuI- and HindIII-cut pWR34 are used in step (e), and primer I (for StuI- and HindIII-cut pWR34) is used in step (f).

Progressive shortening of target DNA with nucleases

Duplex DNA can be progressively shortened from the 3' ends by E. coli exonuclease III, using a ratio of 20 units of exonuclease III per µg of DNA at 23°. The single-stranded ends of the DNA can be cleaved with nuclease S1 to produce blunt-ended duplex DNA[10,20]. Figure 4A shows the digestion of HindIII linearized pWR2 DNA (2.2 kb) with exonuclease III, followed by nuclease S1. The size of DNA was reduced to approximately 1.1 kb after 40 minutes of digestion, or approximately 550 (1100/2) base pairs were removed from each end at a rate of 14 bp per minute. Figure 4B shows the digestion of a 2.2 kb human DNA (EcoRI fragment from p27) with BAL31 nuclease at 23°. The size of DNA was reduced to approximately 1.3 kb after 4 minutes of digestion, or approximately 450 bp were removed from each end at a rate of 120 bp per minute. The rate of digestion by both exonuclease III and BAL31 nuclease can be increased appreciably by using a higher ratio of enzyme/DNA end, or at a higher temperature. However, we found that the rate of digestion is more synchronous[20,22] at 23°.

Determination of the size of progressively shortened target DNA cloned in pWR34

After cloning the progressively shortened 2.2 kb human DNA in step (b) of Figure 1 and cloning the DNA fragments in step (e), white colonies were picked for further analysis. DNA from each size class of

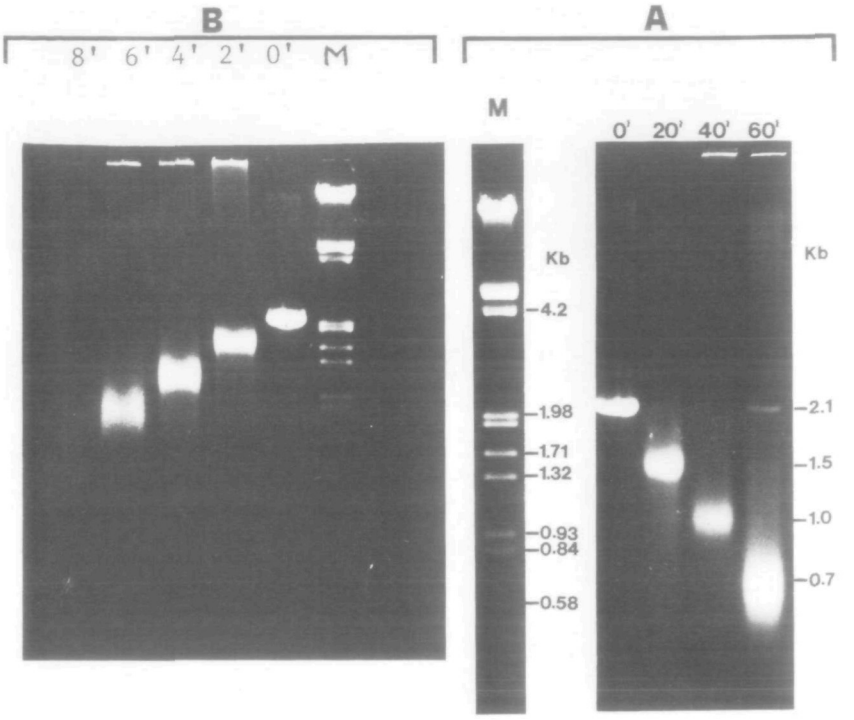



Fig. 4. Progressive shortening of DNA with exonuclease III and nuclease S1 or by nuclease BAL 31, followed by agarose gel electrophoresis. (A) HindIII-linearized pWR2 DNA (36 µg) was digested by 660 units of exonuclease III at 23° (0 min, 20 min, 40 min and 60 min), followed by 240 units of nuclease S1, and digestion for 15 min. Size marker (M) was HindIII-EcoRI digested λ DNA. (B) Digestion of a 2.2 kb human DNA with BAL 31 nuclease at 23° for 0 to 8 min, followed by electrophoresis on an 1% low-melting-point agarose gel. Size marker (M) was HindIII-EcoRI digested λ DNA.

shortened target DNA cloned in pWR34 was isolated by a mini-preparation procedure (see (E) under Method section). The DNA was digested with EcoRI (one of the original site for cloning) and BamHI (a site adjacent to the StuI site for cloning), to cleave out the cloned human DNA from the vector. The samples were fractionated by gel electrophoresis on an 1% agarose gel. In samples 1-13 and sample P of Figure 5, the upper band represents the 2.7 kb vector DNA (pWR34), and the lower band represents the inserted human DNA. The size difference between clones 2 and 3 are about 190 bp. Appropriate clones were then chosen for sequence analysis.

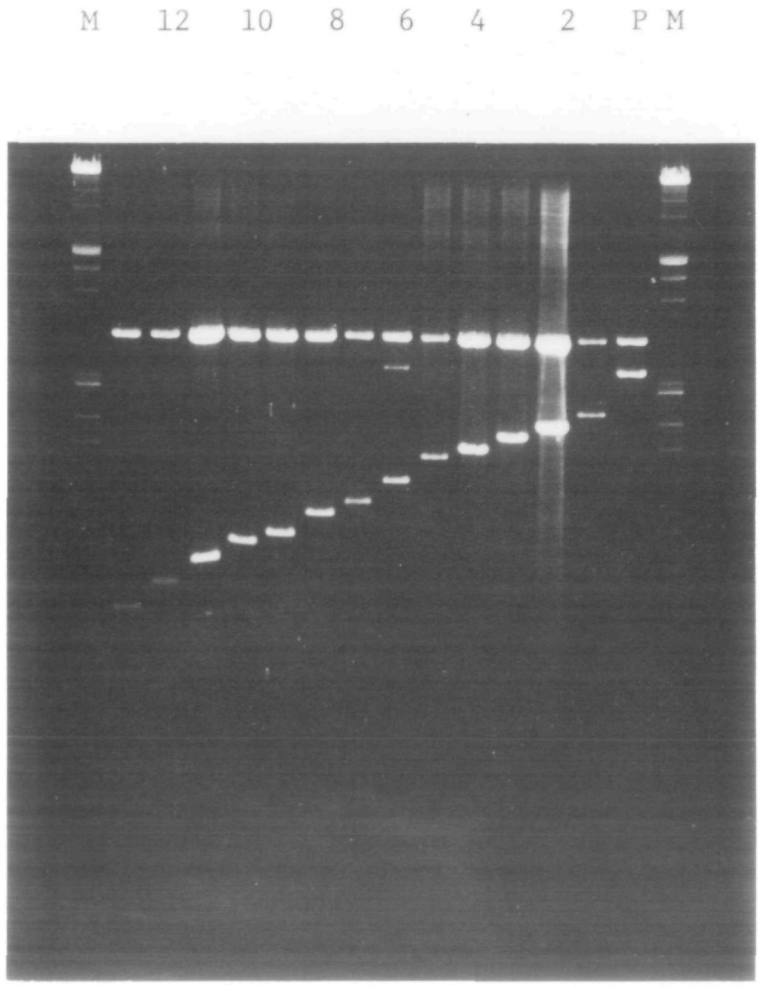



Fig. 5. Progressively shortened human DNA from one given end in different subclones. The 2.2 kb human DNA in HindIII-linearized p27-34 plasmid was progressively shortened from the right-hand end by BAL 31 nuclease, followed by subcloning into pWR34 between StuI and EcoRI sites (see Fig. 1). From each transformation reaction, hundreds of white colonies were obtained. Four of them were selected at random, and the DNA digested with EcoRI and BamHI to liberate the cloned target DNA. The samples were fractionated by gel electrophoresis (1% agarose gel). Size marker (M) was EcoRI-HindIII digested λ DNA. P lane gave EcoRI digested p27-34 DNA. Those colonies with inserts of the desired size were selected for sequence analysis. A portion of the DNA sample from each desired colony was arranged according to size, and a second agarose gel electrophoresis was run to produce the result of this figure.

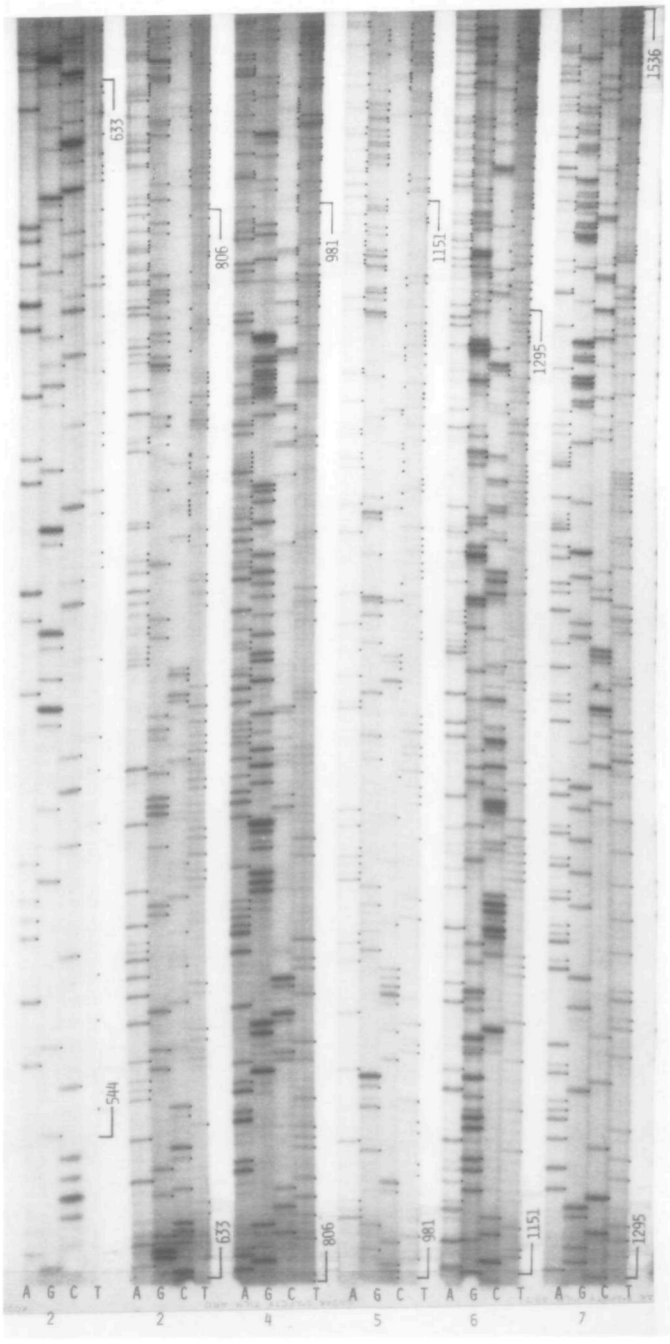
DNA sequence analysis of the human DNA from the subclones

DNA from clones 2,4,5,6 and 7 were chosen for sequence analysis of the first kilo-base of the 2.2 kb human DNA. Each plasmid DNA sample was linearized by digestion with SmaI or PvuII, and heated at 100° for 3 minutes in the presence of primer I (20 times molar ratio over the target DNA). After cooling to 0°C, the dideoxynucleotide chain termination reaction[1,11] was carried out to extend and to label the primer. The reaction mixtures were then fractionated on a 8% polyacrylamide gel. The sequence of approximately 240-280 nucleotides can be read in each set of 4 lanes. Since the clones we chose for sequencing differ in length by approximately 200 nucleotides, the lower half of the 80 cm gel alone gave all the necessary overlapping sequence as shown in Figure 6. Thus, in a single radioautogram, we can read an overlapping sequence of 1 kb.

DNA sequence of the 2.2 kb human DNA

Using the subcloning and sequencing strategy shown in Figure 1, and gels like the one shown in Figure 6, the entire sequence of the upper strand of the 2.2 kb human DNA from the original clone p27 was determined using primer I, and the sequence of nucleotides 1-1800 is shown in Figure 7. A portion of the sequence of this DNA and its complement has been obtained earlier[23] by using the chemical method[2]. Within this sequence of clone p27, the longest open reading frame following an ATG initiation codon can potentially code for two proteins of 52 and 54 amino acids, respectively. One of them is located between nucleotides 1643-1798, the other one is located between nucleotides 1322-1160 of the complementary strand.

The sequence in the clone p27 includes a well known 300 bp Alu family sequence between nucleotides 1417 and 1693. Thus, this fragment of human DNA is a member of the Alu family. This sequence shares only 77% homology with the consensus Alu sequence, which is lower than the average of 88% homology found in the ten clones reported by Deininger et al[24]. Similar to other Alu family sequences, the 300 bp long Alu sequence in clone p27 can be subdivided into two approximately 145 bp repeats. At the end of the 300 bp sequence in p27 there is a stretch of 12 dAMP residues, and the entire Alu sequence is flanked by a 13 bp direct repeat, d(A-A-A-A-A-G-C-A-C-A-T-C-A). Thus, the Alu sequence in clone p27 conforms to the general structure of: direct repeat...Alu sequence...A-rich sequence...direct repeat.



DISCUSSION

Last year, we described two rapid methods for cloning and sequencing a long segment of duplex DNA in plasmids. In the first method[10], the DNA is partially digested with exonuclease III to produce DNA molecules with 3' ends shortened to varying lengths, followed by repair synthesis to extend and label the 3' ends. In this method, any duplex DNA shorter than 900 base pairs can be sequenced from both ends without the need of added primers. For a DNA fragment longer than 1.0 kb, a second method[16,17] can be used. This method is based on the digestion of the target DNA from both ends by exonuclease III for different lengths of time, followed by nuclease S1 digestion to produce a family of progressively shortened duplex DNA, which are blunt-end ligated to HincII-cut pWR2 DNA, and sequenced by using primer I or II. There are two limitations to this method: (1) only 50% of the sequence of each strand of the DNA can be obtained and (2) blunt-end ligation of both ends of the target DNA to the vector is not efficient. The improved method described in this manuscript overcomes both limitations by: (1) shortening the target DNA from one end at a time so that both strands of the DNA from the same plasmid can be sequenced completely, and (2) by cleaving the progressively shortened target DNA with a cohesive-end producing restriction enzyme to produce a cohesive end, which appreciably improves the efficiency of ligating the target DNA to the vector.

For many workers, it is easier to learn to manipulate plasmids than M13 phage. For this reason, cloning and sequencing using plasmids are probably easier than working with M13 phage. Also, the procedure for the selection of the desired clones is simpler and faster in our plasmid method than in the M13 systems[13-15]. Furthermore, some authors have

Fig. 6. A kilo-base sequencing gel. Five clones (2,4,5,6, and 7 from Fig. 5) were sequenced and electrophoresed on the same gel. One μg of DNA from each clone was used and the DNA was linearized by PvuII digestion. The gel (80 cm x 34 cm x 0.04 cm) was 8% polyacrylamide in 8 M urea. There were two loadings for clone 2, and one loading each for the other clones. The voltage was 2000 and 2800, respectively, for the first and second ten hours of electrophoresis. After drying the gel and exposing the X-ray film overnight, a consecutive sequence of 1000 nucleotides (from number 544 to 1530) can be read from one film (35 x 43 cm) exposed on the lower part of the gel. The sequence of about 70 nucleotides overlaps that of the adjacent lane. The first nucleotide in the overlapping region of each clone is numbered on the film, and the sequence corresponds to the numbers on Figure 7.

```
        10          20          30          40          50
TCATCAGGAT TGTTTATATT GCAGCATTGC AGTATGTGAC GTTAAGAAGA

        60          70          80          90          100
GCTACAAGAT AATTTAAAAA AGTAAAAGAA AAAGAAAGCA CCACCTTTAG

        110         120         130         140         150
AATATCCAAG AGAACCACAC CCACTCTAAT CTTTTGGCTA CCAATTTACA

        160         170         180         190         200
TCCCTCTCCA AGCCTCAATG TTCTCATCTA TAAAATGGGA ATATAGTATC

        210         220         230         240         250
CATCAAAGTG GGTTACTGCG AGGATGTGAA AATGCCCTGT AAACCTTAAA

        260         270         280         290         300
ACAATTTATA AATGAGATTA TTGCTTTTTG TAGAATACTT TTTTATGTGA

        310         320         330         340         350
AGCCCTCTAC TTCCCTTATC TCATTTTACA CATACATTCA TTCTTCAAGA

        360         370         380         390         400
ATTTAAAAAT ACATACACAG CTACCATTTG TTAGAGTACT TAGACTATGC

        410         420         430         440         450
AAGATCCATA TTTTGCTATA TTTGACCCTT GCACTGGCCC TATGAGGCAG

        460         470         480         490         500
GTGCCATTCC CCACCTCACC TGTCCCATAT TTCAGTTGAA GAGGAAGCTC

        510         520         530         540         550
AATGAAAGGT TAAGGAATAC AACCAAGGTC AGTAATAAGC AGAGTCAGTA

        560         570         580         590         600
TCAAAGACTG CTCTGGAGTG GACACTTGGT CTAGATTCTG AGTGCTGATC

        610         620         630         640         650
ACTTGACTAC AGTGGCTTTC TGCTGCTACT TGCCAGGAGC CTTGCAGCCA

        660         670         680         690         700
TCCAAAGAGT TAGCATACAA ACAGGCATTC TGTTAGGGTT GACTTTGTGT

        710         720         730         740         750
CCTCCCAAAA GAAGATATAT TAGAATCCTA ACCCCTGCTT CCTCAGAATG

        760         770         780         790         800
TGATCTTATT TGGAGACAGA GCTTGACAGA GGTTAAGCAA ATTAAAATGA

        810         820         830         840         850
AGTCATTAGA ATCGTCTCTA ATTTAATAAG ACTGGCATCC TTATAAAAAG

        860         870         880         890         900
GGATGTGGAC ACAGAGAGAA AAGCATAGAA GGAAGATGAT GTGAGAAGAC

        910         920         930         940         950
ATAGAAAGAG GACAAAAATC TACAAGCCAA GGAGGGAGGC CTGGATCACA

        960         970         980         990         1000
TCCTTTTCTC ATAGCCCTTA GAAGAATCCA CCCTGCTGAA ACCTTGATTG

        1010        1020        1030        1040        1050
TGGACTTCTA CCCCCAGAAC TGTGAAAAAT ACATTTCTGT TGTTTAAGCC

        1060        1070        1080        1090        1100
ACCCAGTTAG TGCGACTTTG TTTATGGCAG CTTTCACAAC CTAATGCCTA

        1110        1120        1130        1140        1150
TTCTGAATCC TCTTTTCTGG ATGAGGCTCT GGAAGAGGAA GAGATTAGGT
```

```
      1160       1170       1180       1190       1200
GACTTACTCA AGAGGTAGGT GAGATGGCTG AGATTTGCCA CCCCATTTTC

      1210       1220       1230       1240       1250
TGTCCACGTG CTGCACATTC ATTGCTAAAA GAAAGGCTAC TCTGGAGCAG

      1260       1270       1280       1290       1300
GTATTTCTTT TGGTTGTATT CACTGTGGTC CCTGGCATTT ATATCAGTGT

      1310       1320       1330       1340       1350
CTGGCATATC ATAAAGCAAT ACTTGTTGAA TAAATGAATG AATGAATGCA

      1360       1370       1380       1390       1400
GATCTTTCTA CCAACTTACC CAGTGTATCA CAGCTAGATT AAAATGATTT

      1410      ↓1420       1430       1440       1450
TTAAAAAGCA CATCAGGCTG GGTACGGTGG CTCATTCCTG TAATCCCAGC

      1460       1470       1480       1490       1500
ACTTTGGGAG GCCGAGGCAG GTCGATCACT TGAGGTCAGG AGTTCGAGAC

      1510       1520       1530       1540       1550
CAGCCTGACC AACATGTGAA ACCCGTCTCT ACTAAAAATT TAGCAGACTG

      1560       1570       1580       1590       1600
GGTGGCAGAC ACCTGTAATC TCAGCTACTT GGGAGGCTGA GGCAGGAGAA

      1610       1620       1630       1640       1650
TCACTTGAGC CTGGGAGGAC GAGGTTGAAG TGAGCTGAGA TCATGCTACT

      1660       1670       1680       1690      ↓1700
GCACTCCAGC CTGGGCAACA AAGTTAGACT CCGCCTAAAA AAAAAAAAGC

      1710       1720       1730       1740       1750
ACATCATCCT GGAACAAGAA TTTCAACAAC TTACAGTCAC TGTCAGAACA

      1760       1770       1780       1790       1800
AAGAGTTTCA TACTTTGTAT TCAACTTCTT CCCACAATCA AAGACTTATA
```

Fig. 7. Nucleotide sequence of a 2.2 kb human DNA. The entire upper strand of the 2.2 kb human DNA was determined using primer I. However, only the sequence of nucleotides 1-1800 is shown here. The numbers on the figure correspond to those on the gel in Figure 6. The sequence between two arrows is the Alu family sequence. Underlined nucleotides represent several of the repeated sequences.

observed a certain instability of relatively large inserts in certain M13 strains (e.g. the mp family[7]).

For progressively shortened target DNA, exonuclease III gives a more uniform rate of digestion, relatively independent to the sequence[22]. On the other hand, BAL31 nuclease has a strong sequence dependency so that G-C rich regions are digested much slower than other regions[22]. However, the rate of BAL31 nuclease digestion is almost ten times faster than that with exonuclease III, thus it may be of advantage to use BAL31 nuclease to digest DNA that are longer than 2 kb.

For the method presented in this manuscript, low-melting-point

agarose is used (Fig. 1, step d) to fractionate the progressively shortened DNA, and to select the target DNA shortened to the desired length for ligation with the vector (in step e). We found that some batches of low-melting-point agarose contain inhibitors for either ligation or transformation so that step (e) may be strongly inhibited. The way to overcome this problem is to use a different batch of agarose or to use an elution method to remove the agarose[26]. Alternatively, we have omitted the gel electrophoresis step (d) in Figure 1, to select the target DNA of the desired length. Instead, we use an analytical agarose gel and a portion of the digestion mixture after step (c), to show that the extent of shortening the target DNA is working as expected. Then we use the remainder of the digestion mixture after step (c), and ligate it directly to StuI- and EcoRI-digested pWR34. Colony hybridization of the white colonies with the target DNA can be used to positively identify the desired clones.

There are extra bands in our sequence gel, but they rarely interfere with the correct assignment of the sequence. Most extra bands that line up with certain stronger major bands are easy to recognize and to discount. Thus, from the sequence gel, as the one presented in Figure 6, we can correctly read about 98% of the sequence. Some extra bands are also found when using the M13 system or the chemical method. For this reason, it is usually necessary to determine the sequence of the complementary strand in any system, in order to obtain 100% reliable sequence information.

When transforming *E. coli* JM83, JM101, or JM103 with pWR34 carrying the target DNA, there are some problems, especially in cloning long target DNA. This is probably related to the fact that these strains are not *hsdr*$^-$. It is better to use a restriction minus strain such as JM83($r^-$), available from Bethesda Research Laboratory or *E. coli* K12 RRIΔM15[29], which is a derivative of RR1, and has the genotype *leu pro thi strA hsdr*$^-$ *m*$^-$ *lacZ* ΔM15 *F*$'$ *lac* $I^Q$ *Z* ΔM15 *pro*$^+$. The yield of the hybrid plasmid DNA in JM83 or JM101 is very high, and up to 50 µg DNA can be obtained from 1/8 of a petri dish of cells. However, the yield of DNA goes down with repeated passages of the plasmid-containing hosts. One way to overcome this difficulty is to use different aliquots of the original transformant for repeated preparations, instead of using serially transferred cells. Alternatively, the hybrid plasmid can be used to re-transform host cells.

Another advantage of pWR34 is its small size (only 2.7 kb), which would give a high ratio of insert DNA to vector DNA after joining to the insert DNA, and it in turn would make any further manipulation of the insert DNA simpler. Obviously, our method and general approach to cloning genes in pWR34, and direct sequencing of DNA in the same plasmid, can be used in other plasmids such as pUR250 of RUther[29], pUC9 of Messing[28], or any other suitable systems including the RF form of M13 DNA. During the preparation of this manuscript we read the interesting paper of Dente et al[30], who described a new series of plasmids, the pEMBL family, which can be manipulated as a plasmid and yield single-stranded DNA upon infection by a F1 phage.

Acknowledgements

We thank Ulrich RUther for E. coli K12 RRIΔM15, Joachim Messing for E. coli strains JM83, JM101 and JM103, Bethesda Research Laboratory for JM83($r^-$), and J. Yun Tso for helpful discussions. The work was supported by research grant GM29179 from the National Institute of Health, U.S. Public Health Service, and a Hatch grant from the U.S. Department of Agriculture, State Agricultural Experiment Station.

+Present address: Department of Pathology, University of Toronto, Banting Institute, 100 College Street, Toronto, Ontario, Canada

*To whom all correspondence should be addressed

REFERENCES

1. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
2. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
3. Gronenborn, B. and Messing, J. (1978) Nature (London) 272, 375-377.
4. Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene 10, 69-73.
5. Anderson, S. (1981) Nucl. Acids Res. 9, 3015-3027.
6. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
7. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acids Res. 9, 309-321.
8. RUther, U., Koenen, M., Otto, K. and MUller-Hill, B. (1981) Nucl. Acids Res. 9, 4087-4098.
9. Wallace, R.B., Johnson, M.J., Suggs, S.V., Miyoshi, K., Bhatt, R. and Itakura, K. (1981) Gene 16, 21-26.
10. Guo, L.-H and Wu, R. (1982) Nucl. Acids Res. 10, 2065-2084.
11. Guo, L.-H and Wu, R. (1983) Methods in Enzymol. 100, 60-96.

12. Hartley, J., Chen, K.K. and Donelson, J.E. (1982) Nucl. Acids Res. 10, 4009-4025.
13. Hong, G.F. (1982) J. Mol. Biol. 158, 539-549.
14. Poncz, M., Solowiejczyk, D., Ballentine, M., Schwartz, F. and Surrey, S. (1982) Proc. Natl. Acad. Sci. USA 79, 4298-4302.
15. Barnes, W.M. and Bevan, M. (1983) Nucl. Acids Res. 11, 349-368.
16. Guo, L.-H. and Wu, R. Symposium, Tapiei, August 1982, Raden Press, in press.
17. Wu, R., Guo, L.-H., Georges, F. and Narang, S.A. (1983) Trans. New York Acad. Sci., Series II, 41, in press.
18. Wu, R. and Taylor, E. (1971) J. Mol. Biol. 57, 491-511.
19. Morrison, D.A. (1979) in Methods in Enzymol. 68, 326-331.
20. Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P. and Tu, C.D. (1976) Biochem. 15, 734-740.
21. Gardner, R.C., Howarth, A.G., Hahn, P., Brown-Lendi, M., Shepherd, R.J. and Messing, J. (1981) Nucl. Acids Res. 9, 2871-2888.
22. Yang, R.C.A., Guo, L.-H. and Wu, R. (1983) in "Frontiers in Biochemical and Biophysical Studies of Proteins and Membranes" (Liu, T.Y., ed.), Elsevier North Holland, in press.
23. Yang, R., Fristensky, B., Deutch, A.H., Huang, R.C., Tan, Y.H., Narang, S.A. and Wu, R. (1983) Gene, in press.
24. Deininger, D.L., Jolly, D.T., Rubin, C.M., Friedman, T. and Schmid, C.W. (1981) J. Mol. Biol. 151, 17-33.
25. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
26. Yang, R.C.A., Lis, J. and Wu, R. (1979) Methods in Enzymol. 68, 176-182.
27. Wu, R. (1976) Methods in Cancer Res. 12, 87-176.
28. Vieisa, J. and Messing, J.G. (1982) Gene 19, 259-268.
29. RUther, U. (1982) Nucl. Acids Res. 10, 5765-5772.
30. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res. 11, 1645-1655.