J. Mol. Biol. (1980) **143**, 161–178

Cloning in Single-stranded Bacteriophage as an Aid to Rapid DNA Sequencing

F. Sanger, A. R. Coulson, B. G. Barrell, A. J. H. Smith and B. A. Roe†

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

(Received 24 March 1980)

An approach to DNA sequencing using chain-terminating inhibitors (Sanger *et al.*, 1977) combined with cloning of small fragments of DNA in a single-stranded DNA bacteriophage is described. Random fragments from restriction enzyme digestion of the DNA are inserted into the *Eco*RI site of the modified bacteriophage M13mp2 (Gronenborn & Messing, 1978) using a linker oligonucleotide. Individual recombinant plaques are collected, 1-ml cultures grown, and the DNA isolated. A "flanking-primer" from the vector is used to determine a nucleotide sequence in each inserted DNA fragment by the chain-terminating method. This is a relatively rapid and simple method of accumulating sequence data. The 2771-nucleotide sequence of the largest *Mbo*1 restriction enzyme fragment from human mitochondrial DNA was determined by this method.

1. Introduction

We have described a rapid and simple method for determining nucleotide sequences in DNA using DNA polymerase and specific chain-terminating inhibitors (Sanger *et al.*, 1977). The method requires single-stranded DNA to act as template for the DNA polymerase. It is therefore very suitable to use with DNAs, such as certain bacteriophages, that occur naturally in the single-stranded form, or for double-stranded DNAs in which the strands can be separated. However, strand separation is sometimes difficult and this limits the scope of the procedure. One method by which this problem can be overcome has been described by Smith (1979). It involves the treatment of linear DNA with a double strand-specific exonuclease, such as exonuclease III. The resultant product, which is derived from the different ends of the two strands, is largely single-stranded DNA and can be used as a template for the sequencing method.

An alternative way of preparing single-stranded templates is by cloning the DNA as a recombinant in a single-stranded bacteriophage. A number of suitable vectors have been described (Barnes, 1978,1979; Gronenborn & Messing, 1978; Herrmann *et al.*, 1978; Ray & Kook, 1978; Boeke *et al.*, 1979) which are derived from the small single-stranded filamentous bacteriophages. The size of these bacteriophage DNAs is not limited by their packaging so that fragments can be inserted and the virus is

† Present address: Department of Chemistry, Kent State University, Kent, Ohio 44242, U.S.A.

0022–2836/80/300161–18 $02.00/ 0

secreted by the host at high titres into the culture medium, which makes their isolation relatively easy. This paper describes a general approach to DNA sequencing using one of these vectors in conjunction with the terminator sequencing method, and its application to a fragment of human mtDNA†.

2. Principle of the Method

The vector we used had been prepared by inserting a fragment of the *lac* operon containing the β-galactosidase gene into a bacteriophage M13 (Messing *et al.*, 1977) and then introducing a single site for the restriction enzyme *Eco*RI into this gene (Gronenborn & Messing, 1978). This recombinant (M13mp2), which grows normally like M13, can be used as a vector to insert *Eco*RI fragments into the single *Eco*RI site. The M13mp2 containing the β-galactosidase gene forms blue plaques in the presence of isopropyl-β-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-β-D-galactoside, but if DNA is inserted into the RI site β-galactosidase is no longer made and thus recombinants can be recognised as "white" plaques. By inserting a second *Eco*RI site in the vector, Heidecker *et al.* (1980) were able to prepare a 96-long DNA fragment that could be used as a flanking primer for sequencing into the end of an inserted DNA. Schreier & Cortese (1979) have described a simple procedure for isolating small amounts of the recombinant DNA and its use as a template for sequencing.

Our general approach is to prepare various restriction enzyme digests of the DNA to be sequenced and to insert the fragments into the M13mp2 vector using an RI linker (G-G-A-A-T-T-C-C: Heyneker *et al.*, 1976; Scheller *et al.*, 1977). White plaques are then collected at random, grown in 1 ml of culture medium and the DNA isolated. A sequence of about 200 nucleotides on each recombinant is then determined by the dideoxy method using the above flanking primer, as described by Schreier & Cortese (1979). Thus the same primer is used on different templates, whereas in the standard dideoxy method different primers are used on one template. Preliminary studies with bacteriophage lambda DNA showed this to be a rapid method of collecting sequence data, especially at the early stages of an investigation where no sequences are yet known.

3. Materials and Methods

The DNA polymerase used throughout this work was DNA polymerase I, Klenow subfragment, obtained either from Boehringer Mannheim or as a gift from A. R. MacLeod. This latter was prepared from a *λpol*A transducing phage (Kelley *et al.*, 1977) and contained approximately 4 units/ml. Restriction enzyme *Fnu*EI was a gift from M. Smith. Other restriction enzymes were obtained from New England Biolabs.

The dideoxynucleotide triphosphates were obtained from Collaborative Research. (α-^{32}P)-labelled deoxynucleotide triphosphates (spec. act. approx. 400 mCi/μmol) were from the Radiochemical Centre, Amersham. Phenol was redistilled, saturated with water and stored in 2 to 5 ml portions at −20°C until used. Small volumes below 1 μl were normally measured using a washed 5 μl graduated Micropet.

Appropriate manipulations involving recombinant bacteriophages were carried out under category II containment conditions as advised by the Genetic Manipulation Advisory Group.

† Abbreviations used: mtDNA, mitochondrial DNA; ddTTP, dideoxythymidine triphosphate.

(a) *Preparation of the DNA*

*Mbo*I restriction enzyme fragments of human placental mtDNA were cloned in the *Bam*HI site of pBR322 in χ1776 (Drouin, 1980). The recombinant plasmid (pmt18) containing the largest *Mbo*I fragment (M1) was transfected into *Escherichia coli* K12 strain Hb101 (Boyer & Roulland-Dussiox, 1969). 3 l of overnight culture were obtained from growth of a single colony of *E. coli* HB101 carrying the pmt18 plasmid. From this 250 μg of pmt18 were isolated by the cleared lysate method (Katz *et al.*, 1973), followed by ethidium bromide/caesium chloride centrifugation (Radloff *et al.*, 1967).

The M1 fragment was cut out from the purified pmt18 recombinant plasmid by treatment with either *Fnu*EI or *Sau*3A, resolved by fractionation on RPC5 at neutral pH (Eshaghpour & Crothers, 1978) and analysed by electrophoresis through a 1% (w/v) agarose horizontal slab gel (Fig. 1). Fractions containing the M1 fragment were pooled, precipitated with ethanol and a portion restricted with *Hpa*II. The digest was resolved on RPC-5 at neutral pH and analysed by electrophoresis through a 2% agarose horizontal gel slab.

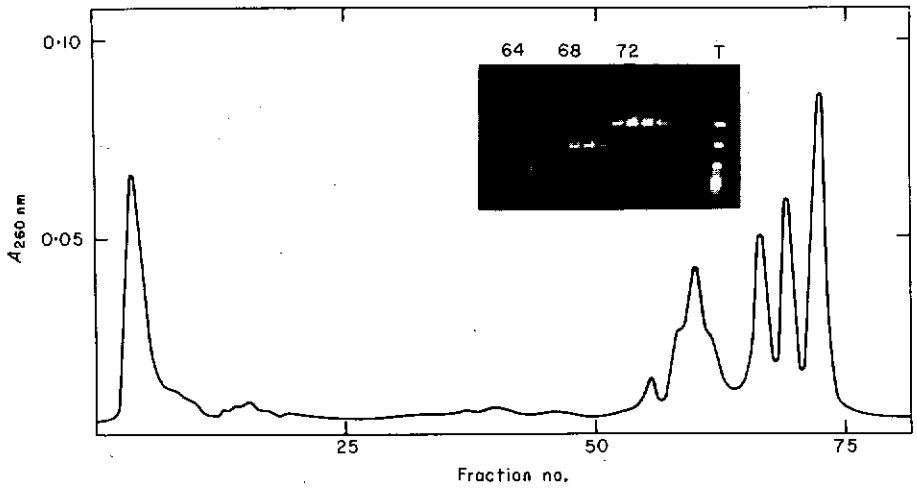


FIG. 1. Fractionation of an *Fnu*EI digest of pmt18 DNA by RPC-5 chromatography. The insert shows a 1% agarose gel analysis of fractions across the gradient. Electrophoresis was from top to bottom. Lane T is a total *Fnu*EI digest of pmt18. The slowest peak on the column and the slowest band on the gel contained fragment M1.

An alternative method for preparing the fragment was by electrophoresis on a 1% low melting agarose gel (Bethesda Research Laboratories). 10 μg of the digested plasmid DNA was applied on a 2 cm wide slot, the gel was stained with ethidium bromide (1 μg/ml) and the band corresponding to the insert visualised under ultraviolet light and excised from the gel. The agarose was melted at 70°C, an equal volume of 10 mM-Tris·HCl (pH 7·6), 0·1 mM-EDTA added and the solution then extracted with the same volume of phenol. After brief centrifugation, the aqueous layer was removed and the phenol extraction procedure repeated a further 2 to 3 times. Finally the solution was extracted with ether and the DNA precipitated overnight at −20°C by the addition of 1/10 vol. 3 M-sodium acetate (pH 5·5) and 2·5 vol. 95% ethanol, followed by centrifugation at 45,000 revs/min for 1 h in a SW60 rotor. The precipitate was washed with 95% ethanol and recentrifuged at the same speed and for the same time. The DNA precipitate was finally dried and redissolved in 10 μl water.

(b) *Digestion and labelling of fragments*

Digestion of the DNA was carried out with various restriction enzymes that recognise a

sequence of 4 nucleotides. In these experiments the enzymes *Alu*I (A), *Hae*III (Z), *Hin*fI (F) and *Hpa*II (Y) were used. (The letters in parentheses refer to the symbol used as a prefix to describe the clones derived from the respective digests.) *Alu*I and *Hae*III cut the DNA to give "flush ends" that can be directly ligated to the linker. *Hin*fI and *Hpa*II give staggered ends which are filled in with DNA polymerase. The fragments were labelled with ^{32}P to facilitate purification on the Sepharose column and for the acrylamide gel electrophoreses that were used to test portions for ligation and *Eco*RI digestion (see Fig. 2). In the case of the *Hin*fI and *Hpa*II digests this was done by including a ^{32}P-labelled deoxynucleotide triphosphate in the DNA polymerase reaction. With the other 2 digests the 3′ termini were labelled by an exchange reaction using DNA polymerase and a labelled triphosphate. In this case the labelling was less efficient. To illustrate the method, the digest with *Alu*I and labelling will be described.

1 μg DNA was treated with 1 to 2 units *Alu*I in 5 μl of solution containing H buffer (6·6 mM-Tris·HCl (pH 7·4), 6·6 mM-$MgCl_2$, 6·6 mM-dithiothreitol and 0·05 M-NaCl) for 1 h at 37°C. 2 μl (2 μCi) [α-^{32}P]dGTP was dried down in a small siliconized test tube and then dissolved in the above solution. 0·5 μl of a solution prepared from equal volumes of 10 mM-dCTP, TTP and dATP was added followed by 0·25 μl DNA polymerase (0·25 unit). Incubation was for 10 min at room temperature. The reactions were terminated by the addition of 1 μl 0·2 M-EDTA (pH 7·4), 20 μl water and 20 μl phenol and vortexing. After standing to separate the layers, the aqueous layer was sucked off using a drawn-out capillary tube and the phenol layer washed with a further 10 μl water. These combined aqueous solutions were extracted twice with about 1 ml ether, and the excess ether evaporated in a stream of air. The solution was transferred to a small siliconized glass centrifuge tube (5 cm × 0·6 cm) and adjusted to 50 μl. 5 μl 3 M-sodium acetate and 125 μl ethanol were added. The solution was left at −20°C overnight and the precipitated DNA collected in an Eppendorf centrifuge (10 min), washed with 0·5 ml ethanol and dried in a stream of air.

The *Hae*III digest was treated as described above. The *Hin*fI digest was labelled by adding it to the dried down residue from 2 μCi [α-^{32}P]dATP. It was then incubated with 0·25 μl DNA polymerase for 5 min. 0·5 μl of a mixture made from equal volumes of the 4 deoxynucleotide triphosphates was added and the incubation continued for a further 20 min. The DNA was then isolated as described above. The *Hpa*II digest was treated as the *Hin*fI digest, except that labelling was with [α-^{32}P]dCTP.

(c) *Ligation with linker*

The synthetic *Eco*RI linker (G-G-A-A-T-T-C-C; Collaborative Research) was not phosphorylated on the 5′ end so was first treated with T_4 polynucleotide kinase (5 units/μl; P. L. Biochemicals). 1 μg of the linker in 9 μl of a solution containing 0·05 M-Tris·HCl (pH 7·4), 0·01 M-$MgCl_2$, 0·01 M-dithiothreitol and 0·5 mM-rATP was treated with 1 μl kinase and incubated for 1 h at 37°C.

The dried ethanol precipitate of the digested and labelled DNA was dissolved in 2 μl of the above phosphorylated linker solution using a drawn-out capillary tube. 0·5 μl of a solution containing 2 mM-rATP, 0·7 mM-spermidine and 0·1 unit of T_4 ligase (Miles) was added. The mixture was incubated in the sealed capillary tube at 14°C overnight and the enzyme inactivated by heating at 70°C for 10 min.

In order to test the efficiency of ligation, a sample (usually 0·1 vol., taken after dilution with the *Eco*RI buffer (see below)) was subjected to electrophoresis on acrylamide gel (in the absence of urea) in parallel with a sample of the unligated digest. If the ligation was successful, the bands in the digest were replaced with much larger products, which were usually unresolved (Fig. 2). Considerable differences were found with different preparations of T_4 ligase. In our experience it was necessary to test several before a satisfactory one was found. It appears that the presence of spermidine may be essential for blunt-end ligation (M. Boguski & W. M. Barnes, personal communication).

(d) *Digestion with* Eco*RI*

15 μl of a solution containing 80 mM-Tris·HCl (pH 7·5), 80 mM-NaCl, 10 mM-$MgCl_2$ and

1 mM-dithiothreitol was added to the above ligation reaction mixture, followed by 3 μl *Eco*RI (5000 units/ml; New England Biolabs). Incubation was for 3 h at 37°C. 1 μl EDTA (pH 7·4), 20 μl water and 20 μl phenol were added. After equilibration, the aqueous layer was removed and the phenol layer washed with a further 10 μl water. The combined aqueous solutions were then extracted with 1 ml ether to remove excess phenol and a small portion run on the acrylamide gel to test for complete RI digestion (see Fig. 2).

The *Eco*RI digest contained a high concentration of small fragments derived from the linker, which were readily ligated into the vector giving white plaques. These fragments were removed on a gel filtration column of Sepharose 4B, which also removed other small fragments of less than about 100 nucleotides from the digest. Small fragments appear to be incorporated into the vector more readily than larger ones, and as they give less sequence information it was an advantage to remove them.

A column of Sepharose 4B (Pharmacia) was prepared in a 1 ml disposable plastic pipette and washed well with 2 mM-Tris·HCl (pH 7·4), 0·1 mM-EDTA. The above *Eco*RI digest was applied to the column, which was monitored for radioactivity using a hand monitor, and 1-drop fractions were collected. There was usually a peak of radioactivity at the "breakthrough", followed by a long tail at a lower activity. 2 to 3 drops at the breakthrough were collected, taken to dryness in a desiccator and dissolved in water (5 μl/drop).

(e) *Incorporation into vector*

The vector used was M13mp2 described by Gronenborn & Messing (1978). The replicative form was isolated and digested with *Eco*RI. 0·5 μl (2 ng) of the vector solution was mixed with 1·2 μl of a solution containing 0·1 M-Tris·HCl (pH 7·5), 20 mM-$MgCl_2$, 10 mM-dithiothreitol and 0·4 mM-rATP. 0·5 μl of the DNA fragment with linkers attached, prepared as described above, was added, followed by 0·25 μl T_4 DNA ligase (25 units/ml). It was incubated at 15°C for 2 h and diluted with 20 μl 10 mM-Tris·HCl (pH 7·4), 0·1 mM-EDTA. Competent cells were prepared from *E. coli* strain JM101 (a *tra*$^-$ mutant, Messing, 1979). The above ligation mixture was diluted into 20 μl 10 mM-Tris·HCl (pH 7·6), 0·1 mM-EDTA and added to 0·2 ml of competent cells, allowed to stand at 0°C for 40 min, heated at 45°C for 3 min and then plated out on 2 plates. The transfection yielded about 100 to 150 plaques, 25 to 30% of which were white (i.e. recombinants).

(f) *Preparation of the cloned DNA*

0·5 ml of a log culture of *E. coli* strain JM101 was added to 50 ml 2 × TY medium (1·6% tryptone, 1% yeast extract, 0·5% NaCl). 1 ml portions were pipetted into sterile glass tubes (10 cm × 1·3 cm) and the white plaques were transferred with tooth picks into the medium. It was found convenient to work up about 36 plaques at a time. The tubes were shaken at 37°C for 7 h and the solutions then transferred to capped Eppendorf tubes and centrifuged for 10 min to pellet the bacteria. The supernatant containing the bacteriophage was transferred to another centrifuge tube, but to avoid dislodging any of the precipitate no attempt was made to transfer it completely. 0·2 ml of a solution containing 2·5 M-NaCl and 20% polyethyleneglycol (M_r 6000) was added and mixed by inverting the tube, which was then allowed to stand at room temperature for about 15 min and centrifuged for 10 min. The supernatant was poured off, taking care to get rid of the last drop without dislodging the small precipitate, which could usually be seen in the tube. After draining the tube, the inside of it was wiped with a Kleenex tissue and 0·1 ml 10 mM-Tris·HCl (pH 7·4), 0·1 mM-EDTA and 50 μl phenol were added. After vortexing and centrifuging for 1 min, the aqueous solution was sucked off using a drawn-out capillary, care being taken not to include any of the phenol layer. It was transferred to another centrifuge tube, 10 μl 3 M-sodium acetate and 0·2 ml cold ethanol were added and the mixture was left at −20°C overnight. The precipitate was collected by centrifugation for 10 min and washed with 1 ml ethanol. The tubes were allowed to drain and any residual ethanol blown off in a stream of compressed air. The DNA was dissolved in 50 μl 10 mM-Tris·HCl (pH 7·4), 0·1 mM-EDTA. This solution was used for sequencing with the flanking primer as described below.

Using the above method, the DNA in the 1 ml cultures from 36 plaques can be worked up in about 3 h. We have normally done this immediately after growing the cultures. If necessary the culture supernatant can be stored in the refrigerator after pelleting the cells; however, in this case it is advisable to centrifuge again before proceeding in case more cells have grown on standing.

(g) *Preparation of the 96-nucleotide primer*

The primer was generated by an *Eco*RI digest of M13mp2962 (Heidecker *et al.*, 1980) and separated by acrylamide gel electrophoresis or RPC-5 high pressure liquid chromatography. For gel separation, 200 μg of *Eco*RI-digested DNA in a volume of 200 μl containing 10% sucrose, 0·03% bromophenol blue and 10 mM-EDTA was loaded onto a 9-cm slot width of an 8% (w/v) polyacrylamide gel (3 mm × 20 cm × 40 cm). Electrophoresis was carried out for 4 h at 300 V in a continuous buffer system containing 0·09 M-Tris-borate (pH 8·3), 2·5 mM-EDTA. The gel was stained with ethidium bromide (1 μg/ml) for 20 min, the DNA visualised with ultraviolet light, and the gel band excised, the gel volume being kept to a minimum. The DNA was eluted by the method of Maxam & Gilbert (1977), precipitated with ethanol as described above, and transferred to a siliconised glass tube, dried and resuspended in 200 μl water before storage at −20°C.

Fractionation on RPC-5 was essentially as for plasmid fragments. The primer DNA eluted from the column at 0·6 M-NaCl.

(h) *Preliminary screening with ddT reaction*

1 μl of the DNA solution prepared as described above was mixed with 1 μl of the "R1-R1" primer and sealed in a drawn-out capillary tube. It was heated in boiling water for 3 min to denature the DNA and annealed at 67°C for about 20 min. To this was added 2 μl of a reaction solution prepared as follows: 10 μl [^{32}P]dATP (10 μCi) was dried down in a small siliconised test tube. The residue was dissolved in 40 μl of a solution containing H buffer, 0·1 mM-dGTP, 0·1 mM-dCTP, 5 μM-TTP and 0·25 mM-ddTTP. 2 μl DNA polymerase (2 units) was added immediately before addition to the above annealed primer-template solution.

Incubation was carried out at room temperature for 15 min. 1 μl 0·5 mM-dATP was added and after a further 15 min the reaction was terminated either by freezing the solution or by the addition of 5 μl deionised formamide containing 0·01 M-EDTA, 0·03% xylene cyanol FF and 0·03% bromophenol blue. This latter solution was usually added just prior to running the electrophoresis gels. Less satisfactory results were obtained if the reaction mixtures were stored at −20°C after adding the formamide solutions. After heating at 100°C for 3 min,

---

FIG. 2. Autoradiograph of a 4% non-denaturing polyacrylamide gel illustrating the preparation of *Taq*I restriction fragments with *Eco*RI "sticky ends" for cloning into the *Eco*RI site of the M13mp2 vector. The gel (40 cm × 20 cm × 0·35 mm) was run in a continuous buffer of Tris-borate, EDTA at a constant current of 20 mA until the bromophenol blue tracking dye had reached the bottom. The *Taq*I digest was of the mtDNA insert of the plasmid pmt1, which contains a $1{\cdot}7 \times 10^3$ base *Bam*I-*Eco*RI fragment and the $1{\cdot}1 \times 10^3$ base *Eco*RI-*Eco*RI fragment (which is also contained within the insert of pmt18). The *Taq*I fragments were blunt-ended and labelled with ^{32}P by extension of their 3′ ends with DNA polymerase and [α-^{32}P]dCTP and dGTP. Track (a) is a portion of this reaction. Track (b) is the same ^{32}P-labelled digest after blunt-end ligation with *Eco*RI linker. Track (c) is the result of an almost complete *Eco*RI digestion of the products of the linker ligation reaction. Bands 1 to 5 in track (c) are interpreted as follows: (1) *Taq*I fragment onto which no *Eco*RI linkers have been ligated. (2) *Taq*I fragment with only a single linker at one of its ends, which has been cleaved with *Eco*RI. (3) *Taq*I fragment with an *Eco*RI-cleaved linker at each end. (This is the predominant product from the reactions.) (4) *Taq*I fragment ligated to 2 *Eco*RI linkers only 1 of which is cleaved with *Eco*RI to give the *Eco*RI sticky end. (5) A band not present in the original *Taq*I digest which is probably the result of 2 of the original fragments of the digest blunt-end ligating together directly or *via* a linker(s) that has not been cleaved during the subsequent *Eco*RI digest. This is supported by the presence in track (b) of a fragment of this size with linkers added.

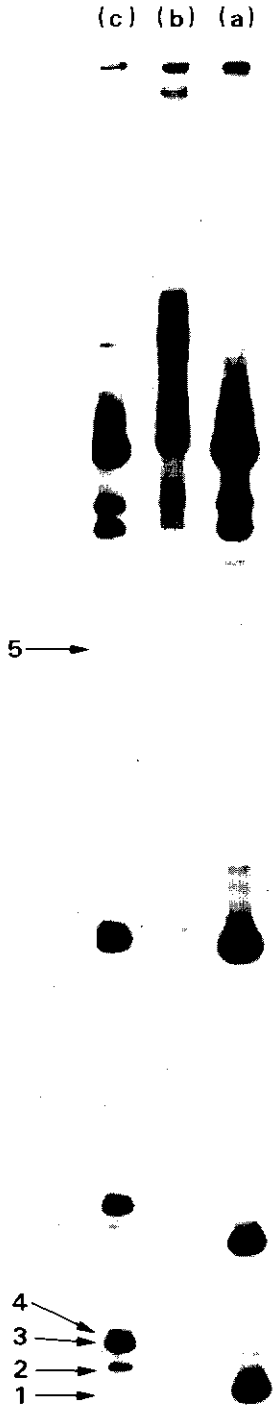


FIG. 2

about half the solution was applied to an 8% acrylamide gel for electrophoresis (Sanger & Coulson, 1978). Usually the reaction was carried out with 18 samples at a time, which were run together on a single gel.

(i) *Sequencing of the DNA*

The sequencing procedure was carried out essentially as described previously (Sanger *et al.*, 1977; Sanger & Coulson, 1978; Schreier & Cortese, 1979) using the 96-nucleotide R1-R1 primer. When this untreated fragment was used to prime, a strong artifact band was present in all 4 channels on the gel at a position about 130 nucleotides from the 5′ end of the primer, and the sequence before this could not be read satisfactorily. This effect is believed to be due to the 3′ end of the strand of the primer, which has the same sequence as the template, looping back to base-pair intramolecularly at a position where it can then act as a primer for DNA polymerase to extend by a further 40 residues. This effect could be avoided by initial treatment of the primer with exonuclease III (Schreier & Cortese, 1979). The sequence of the insert back to and including the *Eco*RI site of the linker could usually then be read unambiguously.

In a normal sequencing experiment 5 μl of the DNA solution from the clone was mixed with 5 μl of the exonuclease-treated primer in H buffer, the mixture heated and annealed, and 2 μl samples used in the 4 reaction mixtures with the appropriate dideoxynucleotide triphosphates, as described by Sanger *et al.* (1977). With different preparations of the dideoxynucleotide triphosphates we found some variation in the concentrations required to give optimum results. The concentration was usually established empirically. Also the concentrations for the G and C reactions had to be varied according to which strand of the mtDNA was being sequenced, as they have a very asymmetric G:C ratio. Normally two 6% acrylamide gel electrophoreses were run with each clone, half of the denatured reaction mixture being applied to each gel. One gel was run at 25 mA for 3 h and the other for 6 h. From the autoradiographs of these 2 gels a sequence of between 150 and 300 nucleotides could usually be determined.

4. Results

(a) *Preparation of the DNA*

In this paper we described the application of this approach to the sequence of a 2771-nucleotide long fragment of human mtDNA. The DNA was prepared from a clone made by Drouin (1980), who subjected human mtDNA to digestion with the restriction enzyme *Mbo*I and inserted the fragments into the *Bam*HI site of the plasmid pBR322. Figure 3 shows the *Mbo*I map of human mtDNA, together with some other features of the DNA that have been mapped. The fragment studied here is the largest *Mbo*I digestion product (fragment M1) and was isolated from clone pmt18 (Drouin, 1980). A certain amount of sequence data was already available on this fragment; in particular, A. J. H. Smith had determined the sequence of the *Eco*RI fragment, which is completely contained within the *Mbo*I fragment and had been isolated from a different clone (pmt1). Other sequence data had also been obtained, both on intact mtDNA and on pmt18 using the exonuclease method (Smith, 1979), but this is not described here.

The DNA of fragment M1 was isolated from the clone by fractionation on an RPC-5 column (Fig. 1). An alternative technique, which has been used with other DNAs, is to use electrophoresis on low melting agarose gels. The latter method is particularly suitable for small amounts. Advantages of the RPC-5 chromatography for the fractionation of DNA restriction enzyme digests are: (a) the fractionation

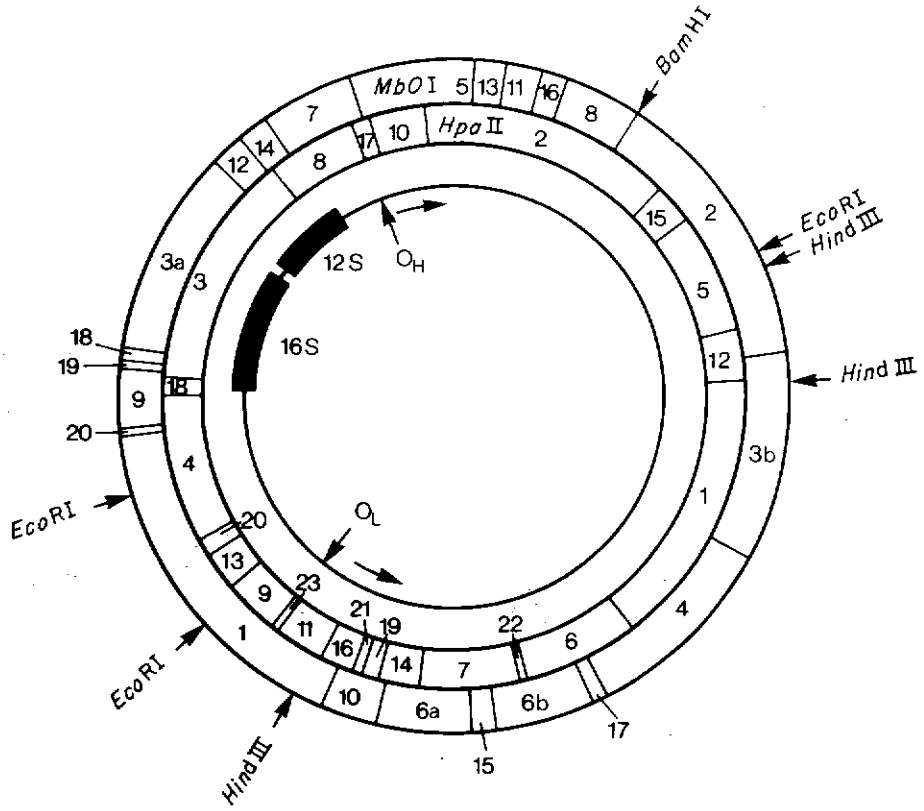


FIG. 3. Restriction enzyme digestion map of human mtDNA. Outer ring *Mbo*I (Drouin, 1980); inner ring *Hpa*II (Ojala & Attardi, 1977). The splitting sites for *Eco*RI, *Bam*HI and *Hin*dIII, the position of the ribosomal RNA genes (12 S and 16 S) and of the origins of heavy strand (O_H) and light strand (O_L) DNA replication are also shown.

step is completed in less than one hour, (b) the desired fragments can be directly precipitated with ethanol or dialysed and lyophilised, and (c) a single column can be re-used several (at least 30) times without loss of resolution. However, the resolution of RPC-5 is greatly reduced when separation of an *Eco*RI digest is attempted: thus the fragments obtained by *Eco*RI digestion of DNAs that generate oligomers similar in size to those shown in Figure 1 are not resolved. In other experiments we observed that RPC-5 can successfully resolve fragments obtained by restriction digestion with *Hin*dII, *Hin*dIII, *Hha*I, *Hae*II and *Hae*III. While these experiments were in progress similar observations were reported by Patient *et al.* (1979).

From the digest prepared with the four restriction enzymes, as described in the Materials and Methods section, we isolated 72 plaques. Cultures (1 ml) of each were grown and the phage isolated from the culture fluid by precipitation with polyethylene glycol, as described. An alternative method of isolation, which gave

equally suitable DNA, was by pelleting the phage in the ultracentrifuge for six hours at 450,000 *g* (Schreier & Cortese, 1979). In general 1 ml cultures gave enough DNA for ten sequencing experiments.

(b) *Preliminary screening of clones*

In order to decide which clones to study, a preliminary screening was carried out by doing the normal sequencing reaction with only one dideoxynucleotide, in this case ddTTP. These ddTTP reaction mixtures were run in parallel on an acrylamide gel. An example of such a gel is shown in Figure 4. A few clones gave no pattern of bands, or were very weak, indicating that there was insufficient DNA present; these were not investigated further. Some others gave an identical pattern of bands, indicating that they contained the same fragment of DNA (e.g. samples Y6 and Y10, Fig. 4). Clones containing small inserts could also be detected by the presence of a band pattern representing the nucleotide sequence in the vector on the 3′ side of the insert. This common pattern can be seen in clones A2, F5 and Y10 (Fig. 4) and the arrows show where the insert is ligated to the vector. From the position of this pattern the size of the insert can be estimated. It was also possible to identify clones that were from a part of the DNA that had already been sequenced. From the pattern of bands it was possible to write the sequences in the forms of Ts and Xs, X representing A, G or C. Thus clone F3 could be written T-T-X-T-X-X-X-X-X-X-T-X-X-X-T-T-X-X-X-T-T-X-T-X-X-T-X-X-T-X-X-T, etc. Usually if the autoradiograph film was over-exposed a background "ladder" would show up on which the strong T bands were superimposed. In this way it was much easier to estimate the number of nucleotides between the T residues, and in the analysis it is in fact more important to know the number of nucleotides than the actual position of all the Ts. The sequence was then checked against the known sequences previously stored in the computer, using the program SEQFIT (Staden, 1977), and it could be decided which clones should be sequenced.

(c) *Sequence of fragment M1*

Figure 5 shows the location of the different cloned fragments on the physical map of the *L*-strand of fragment M1. This also shows which were subjected to the full sequencing procedure and which were only positioned using the ddT screening. For instance, most of the fragments identified as being present between the two *Eco*RI sites (positions 430, 1582) were not studied as this sequence was already known. The fragments from M1 cover the whole of the DNA except for a gap in the region between nucleotides 1582 and 1686, which contains one of the *Eco*RI sites. This site presents a particular problem since it is always digested by the *Eco*RI used in the above procedure and is never found intact in a clone. In order to obtain a clone covering this region, fragment M1 was digested with restriction endonuclease *Hpa*II and the digest fractionated by RPC-5 chromatography. A fraction containing *Hpa*II fragments 9 and 11 (Fig. 5, positions 1154 to 2050 and 2074 to 2570) was then subjected to the random M13mp2 cloning procedure. In order to obtain a sequence spanning the *Eco*RI site, the DNA was methylated with *Eco*RI methylase (Maniatis *et al.*, 1978) which renders it resistant to the restriction enzyme. Recombinants were prepared from the *Hpa*II fragments 9 and 11 without further digestion, and from

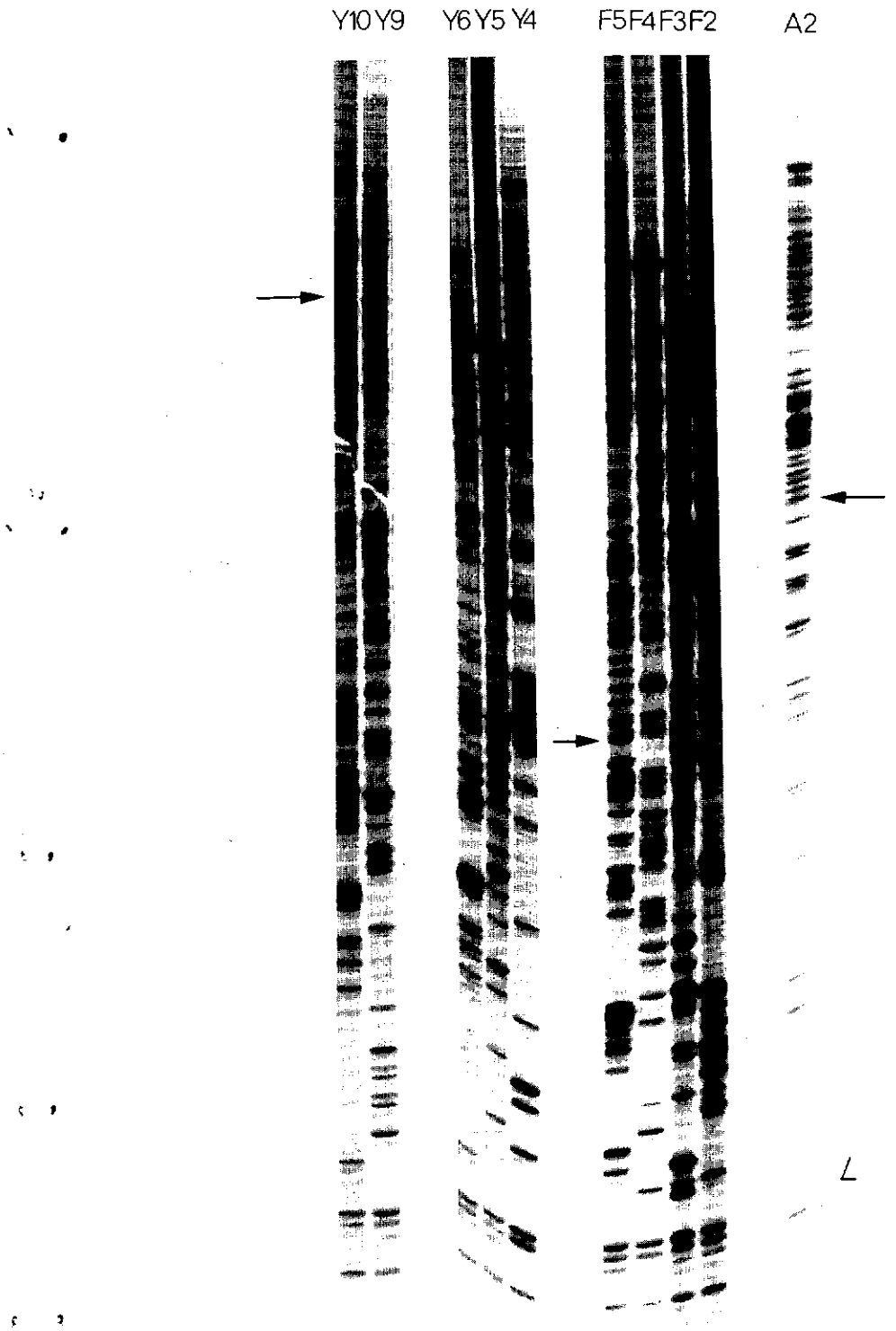


FIG. 4. Preliminary screening of some of the clones using the ddTTP reaction. (For details see text.) The origins of the gels are 5 to 10 cm off the top of the Figure. *L* marks the end of the 96-long R1-R1 primer. The arrows indicate the beginning of the vector sequence on the 3′ side of the insert.

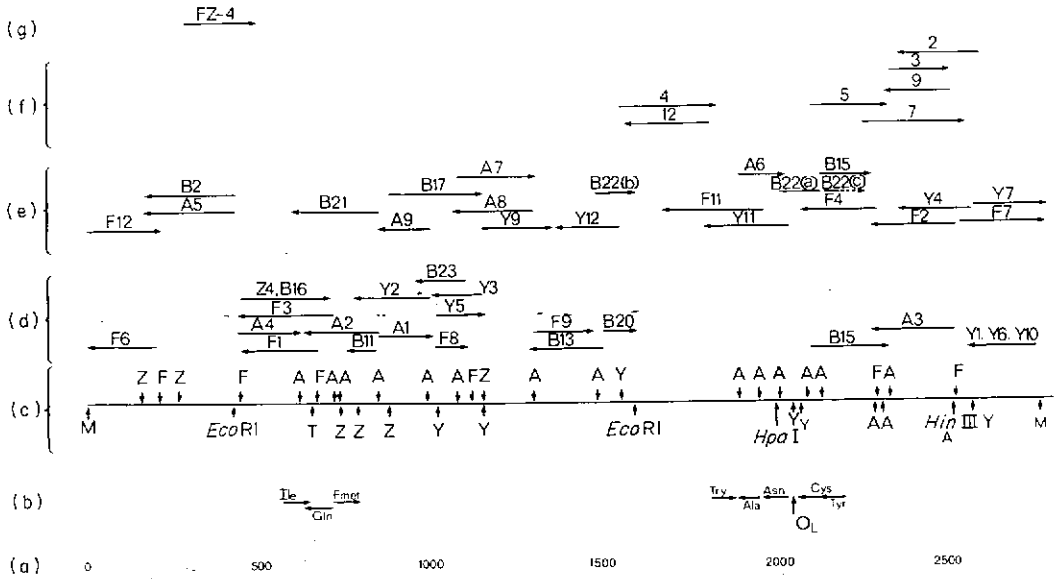


FIG. 5. Map of fragment M1 showing the location of the cloned DNA fragments. (a) Nucleotide number; (b) position of tRNA genes and origin of *L*-strand replication (O_L); (c) restriction enzyme sites, A = *Alu*I, F = *Hin*fI, M = *Mbo*I, Y = *Hpa*II, Z = *Hae*III; (d) location of inserts in clones that were not sequenced. (Their position was determined from the preliminary screening with ddT.) The arrows indicate the position of the 5′ end of the insert and direction of reading in the insert. In the case of the smaller clones the length of the arrow represents the length of the insert, but for the larger ones the length of the insert was not known. The fragments with prefix B were from a mixture of various restriction enzyme digests. (e) Location of inserts in clones that were sequenced. The length of the arrows shows the length of the sequence that was determined. Clone B22 contained 3 separate *Alu*I fragments ((a), (b), (c)) that had presumably been ligated together during the cloning procedure. (f) Clones derived from digests of *Hpa*II fragments 9 and 11 (see text). (g) Position of sequence previously determined by priming with fragment FZ4 on exonuclease III-treated pmt18 (Smith, 1979).

digests of them with restriction endonucleases *Alu*I and *Hin*dII. Two clones (4 and 12) covered the desired sequence and the *Eco*RI site. None of the clones from the whole M1 digests covered the other *Eco*RI site at position 430; however, this region had already been spanned in previous work by priming on the exonuclease-digested pmt18 with restriction enzyme fragment FZ4 (see Fig. 5(g)).

5. Discussion

Although extensive sequences have been determined with the methods using controlled synthesis with DNA polymerase (Sanger & Coulson, 1975; Sanger *et al.*, 1977), it has sometimes been found difficult to obtain satisfactory results with certain DNA preparations. Probably this is usually due to the presence of impurities. Small amounts of contaminating DNA can make the reading of the sequence ambiguous and the activity of DNA polymerase can be adversely affected by other contaminants, both in the template and the primers, so that some care must be taken in their preparation. In the method described here the same primer is used for each clone so that it can be prepared on a relatively large scale in a pure

form, and the use of an RPC-5 column for this avoids the necessity for using gel electrophoresis, which can sometimes introduce inhibiting impurities. Furthermore, the templates prepared from the cloned DNA isolated as described above are usually sufficiently pure to give satisfactory results by the dideoxy method.

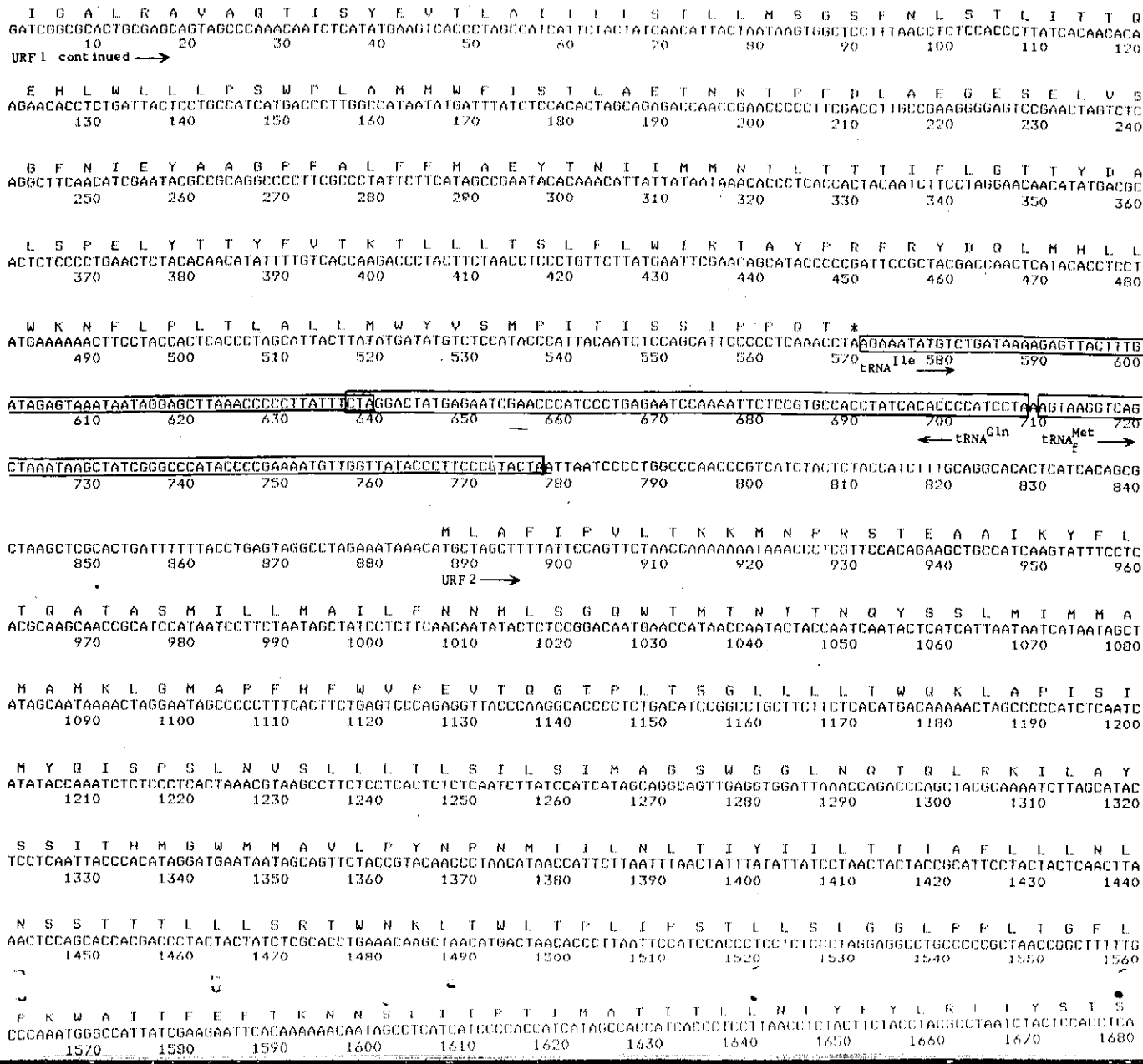
Another important advantage of the method is that it does not involve any physical fractionation of restriction enzyme fragments. In all other rapid DNA sequencing methods this is an essential step that can be quite laborious. In the present method the fractionation is replaced by a cloning procedure which necessarily yields pure DNA, and it can in theory be applied to DNA of any size. With the normal strategies the size of DNA that can be studied is limited by the resolving power of methods for fractionating restriction enzyme fragments. In the present method the only fractionation is by spreading the clones on a plate and the purity of the products is therefore independent of the complexity of the DNA digest. We have recently started to apply the method to bacteriophage λ DNA (approx. 49,000 nucleotides) and have rapidly accumulated data covering about 40,000 nucleotides. In comparison the isolation of pure restriction enzyme fragments from a DNA of this size is a formidable problem involving at least two successive digestion and fractionation steps.

Normal procedures for sequencing are usually preceded by the preparation of a restriction enzyme fragment map as a basis to a systematic approach. In the random method described here this is unnecessary since it aims to determine the whole sequence by overlapping. Because of this, however, it does depend more heavily on the use of computer analysis to interpret and store the data.

Using the random approach of collecting clones, sequence data can be obtained relatively rapidly, but towards the end of an investigation new data will accumulate more slowly along with increasing amounts of redundant data. We have found that much of this redundant data is useful for confirmation and completion of the sequence, but it may be preferable to use a more systematic approach to complete a sequence. In this investigation we have used the digest from a smaller restriction enzyme fragment to complete the sequence. Alternatively one could use hybridization probing to isolate selected clones, or prime with internal primers on the larger templates.

Ideally this method requires a completely random collection of clones in which the different regions of the DNA are equally represented. From Figure 5 it would appear that this is not entirely the case: many clones were found that covered the regions 1 to 1580 and 2070 to 2770 and only a few for region 1580 to 2070. Similarly, Drouin (1980) found that some of the *Mbo*I fragments of human mtDNA were particularly difficult to clone in the plasmid pBR322. How general this phenomenon is, and whether it will constitute a serious drawback to the method, can only be found out by further experience.

The method as described here requires digestion of the linkers with *Eco*RI endonuclease so that any *Eco*RI sites in the DNA are also split and no clones are obtained which contain them intact. This problem can be avoided by preliminary treatment of the DNA with *Eco*RI methylase (Maniatis *et al.*, 1979) which protects the sites from digestion by the endonuclease. This technique was used to obtain a


URF 1 continued
tRNA Ile
tRNA Gln
tRNA Met
URF 2

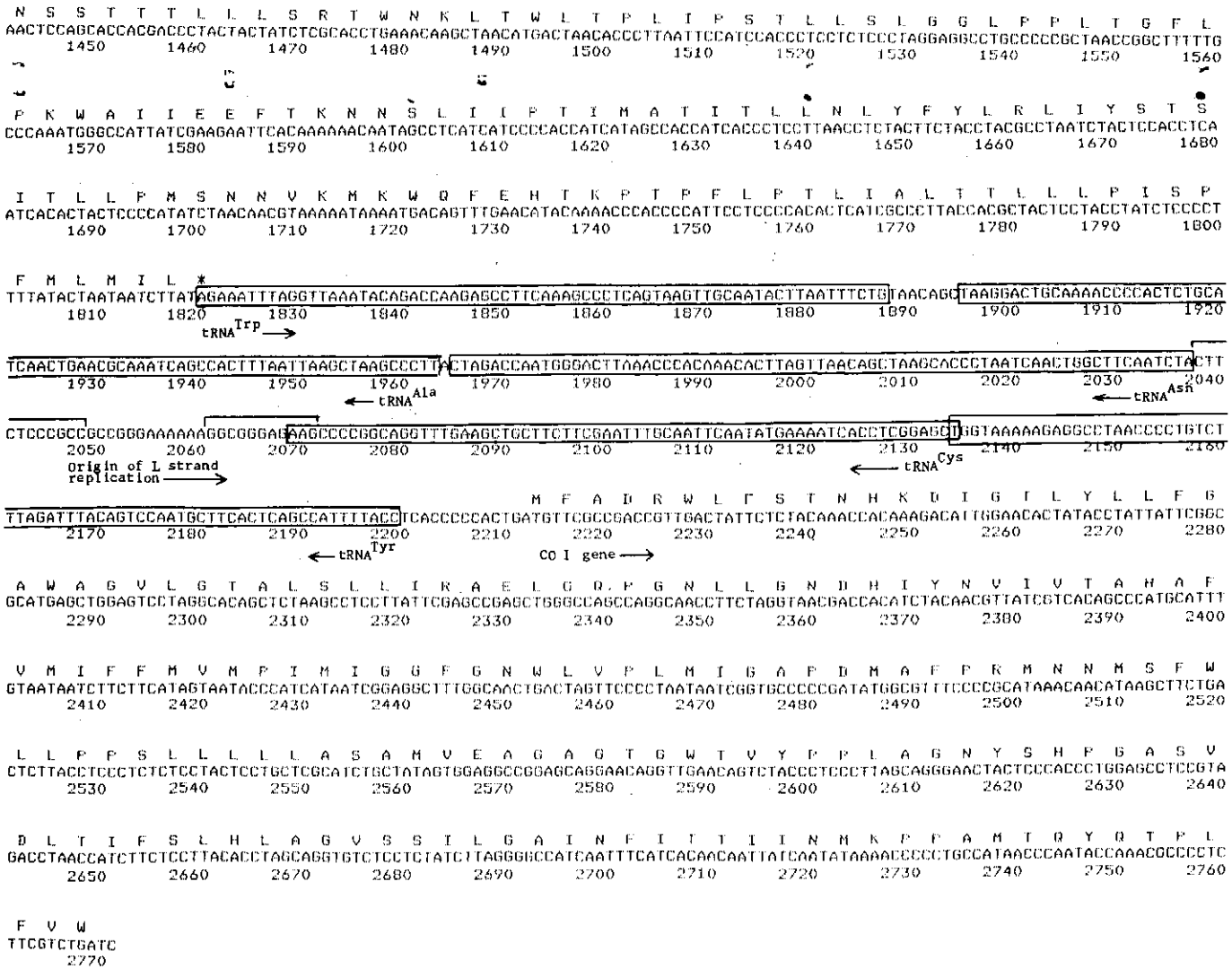


FIG. 6. Nucleotide sequence of fragment M1 from human mtDNA. The positions of the probable tRNA genes are indicated by boxes, the arrows showing the direction of transcription of the RNAs. The probable amino acid sequence of the proteins coded by the DNA is indicated by the single letter code above the corresponding DNA sequence. URF, unidentified reading frame. Overlining indicates the position of a probable hairpin loop that is associated with the origin of *L*-strand replication (Martens & Clayton, 1979). Hyphens omitted for clarity.

sequence spanning the site at position 1582, but it was not used for the other site (position 430) as its sequence was already known.

During the ligation reaction with the linker oligonucleotide restriction enzyme fragments present may be ligated together, leading to the formation of a new bond that was not present in the original DNA being studied. This possibility must be taken into account when interpreting the results. Because of the high concentration of linker it does not occur frequently, the only example in the present work being clone B22 (see Fig. 5), in which three *Alu* fragments were ligated together. (See also Figure 2.) In the case of restriction enzymes that give "flush ends", ligation leads to reformation of the restriction site, which can be recognised. For enzymes giving extended 5′ ends, which are filled in with DNA polymerase, the site is not necessarily restored but a predicted sequence is obtained which can be searched for. For example, *Hinf* (cutting site G-A-N-T-C) will ligate to give G-A-N-T-A-N-T-C. In the case of enzymes giving an extended 3′ end (e.g. *Hha*I or *Hph*I), the problem appears to be more serious. In fact in a single experiment in which these enzymes were used with lambda DNA almost all the clones contained more than one inserted fragment ligated together. The exact explanation of this is not clear but it was probably at least partly due to the difficulty of converting them to flush ends.

It has been observed that small fragments are more readily inserted into the vector M13mp2 than large fragments and that inserted large fragments (over about 2000 nucleotides) are frequently lost from the clones (see, for example, Cordell *et al.*, 1979). Most of the clones used in this method are for small fragments, but it does suggest the possibility that a section of an insert might be deleted and this would lead to a false result. We have not detected this effect, but as the method depends on overlapping sequences and generating a large amount of redundant data this effect would be seen if it occurred.

Figure 6 shows the complete sequence of the DNA fragment. According to the mapping experiments of Angerer *et al.* (1976), the region codes for a number of transfer RNAs. We have therefore looked for tRNA genes, using the program TRNA of Staden (1980) which identified sequences that can be folded in the cloverleaf structure. Eight such structures, arranged in two groups, have been found and identified by the sequence in the anticodon position. As with other human mt-tRNAs (Barrell *et al.*, 1979; Eperon *et al.*, 1980), they are all somewhat anomalous when compared with conventional tRNAs in that some of the so-called "invariable" features are missing. These tRNAs will be discussed in more detail in a later paper in connection with other tRNA structures from human mitochondria. Of particular interest is the sequence identified as $tRNA^{Trp}$. It has the anticodon TCA, which corresponds to the codons UGG and UGA, whereas other $tRNA^{Trp}$s that have been identified have the anticodon CCA, corresponding to UGG only. This is consistent with the identification of TGA as a codon for tryptophan in mitochondria (Barrell *et al.*, 1979; Macino *et al.*, 1979; Fox, 1979) rather than as a termination codon.

Martens & Clayton (1979) have identified the origin of *L*-strand replication (O_L) of mouse DNA and have determined a sequence of 318 nucleotides containing it. Their sequence is homologous to positions 1849 to 2167 of Figure 6. The human sequence also contains a strongly base-paired loop (positions 2038 to 2071) in the region

corresponding to the loop where the origin is located in the mouse mtDNA. Thus we conclude that this is the origin in the human mtDNA. Either side of this loop we have identified tRNA genes (Fig. 6). Homologous structures can also be noted in the sequence from the mouse DNA. Indeed all five tRNA genes that we find in this region of human mtDNA seem to be present in the mouse and to have the same anticodons. As judged by the distribution of the termination codons TAG and TAA, the sequence on either side of and between the two tRNA clusters contains open "reading frames" which probably code for proteins. These also will be discussed in detail in a later paper. J. E. Walker (personal communication) has identified the reading frame starting at position 2212 as the gene for cytochrome oxidase subunit I by comparison of the DNA sequence with the amino acid sequence at the N-terminus of the bovine protein.

We thank Dr A. R. MacLeod for a gift of DNA polymerase, Dr M. Smith for a gift of restriction enzyme *Fnu*EI and Dr G. Attardi for a gift of HeLa cell mtDNA and for helpful advice. We also thank Mr R. Staden, Drs G. P. Winter and P. Schreier for useful advice and discussions in connection with this work, and Drs J. Messing and B. Gronenborn for making available to us their strains of bacteriophage and bacteria and their protocols.

One of the authors (B. A. R.) is a recipient of a National Institutes of Health Research Career Development Award (KO-4-GM-00178) and was on leave from Kent State University.

REFERENCES

Angerer, L., Davidson, N., Murphy, W., Lynch, D. & Attardi, G. (1976). *Cell*, **9**, 81–90.
Barnes, W. M. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4281–4285.
Barnes, W. M. (1979). *Gene*, **5**, 127–139.
Barrell, B. G., Bankier, A. T. & Drouin, J. (1979). *Nature (London)*, **282**, 189–194.
Boeke, J. D., Vovis, G. F. & Zinder, N. D. (1979). *Proc. Nat. Acad. Sci. U.S.A.* **76**, 2699–2702.
Boyer, H. W. & Roulland-Dussiox, D. (1969). *J. Mol. Biol.* **41**, 459–472.
Cordell, B., Bell, G., Tischer, E., DeNoto, F. M., Ullrich, A., Pictet, R., Rutter, W. J. & Goodman, H. M. (1979). *Cell*, **18**, 533–543.
Drouin, J. (1980). *J. Mol. Biol.* **140**, 15–34.
Eperon, I. C., Anderson, A. & Nierlich, D. P. (1980). *Nature (London)*, **286**, 460–467.
Eshaghpour, H. & Crothers, D. M. (1978). *Nucl. Acids Res.* **5**, 13–21.
Fox, T. D. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 6534–6538.
Gronenborn, B. & Messing, J. (1978). *Nature (London)*, **272**, 375–377.
Heidecker, G., Messing, J. & Gronenborn, B. (1980). *Gene*, **10**, 69–73.
Herrmann, R., Neugebauer, K., Schaller, H. & Zentgraf, H. (1978). In *The Single-stranded DNA Phages* (Denhardt, D. T., Dressler, D. & Ray, D. S., eds), pp. 473–476, Cold Spring Harbor Laboratory, New York.
Heyneker, H. L., Shine, J., Goodman, H. M., Boyer, H. W., Rosenberg, J., Dickerson, R. E., Narang, S. A., Itakura, K., Lin, S. & Riggs, A. D. (1976). *Nature (London)*, **263**, 748–752.
Katz, L., Kingsburg, D. K. & Helinski, D. R. (1973). *J. Bacteriol.* **114**, 577.
Kelley, W. S., Chalmers, K. & Murray, N. E. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5632–5636.
Macino, G., Coruzzi, G., Nobrega, F. G., Li, M. & Tzagoloff, A. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 3784–3785.
Maniatis, T., Hardison, R. C., Lacey, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstatiadis, A. (1978). *Cell*, **15**, 687–701.
Martens, P. A. & Clayton, D. A. (1979). *J. Mol. Biol.* **135**, 327–351.
Maxam, A. M. & Gilbert, W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 560–564.
Messing, J. (1979). *Recombinant DNA Technical Bulletin*, **2**, 43–48.

Messing, J., Gronenborn, B., Müller-Hill, B. & Hofschneider, P. H. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 3642–3646.
Ojala, D. & Attardi, G. (1977). *Plasmid*, **1**, 78.
Patient, R. K., Hardies, S. C., Larson, J. E., Inman, R. B., Maquat, L. Y. & Wells, R. D. (1979). *J. Biol. Chem.* **254**, 5548–5554.
Radloff, R., Bauer, W. & Vinogard, J. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 1514–1520.
Ray, D. S. & Kook, K. (1978). *Gene*, **4**, 109–119.
Sanger, F. & Coulson, A. R. (1975). *J. Mol. Biol.* **94**, 441–448.
Sanger, F. & Coulson, A. R. (1978). *FEBS Letters*, **87**, 107–110.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.
Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. & Davidson, E. H. (1977). *Science*, **196**, 197–200.
Schreier, P. H. & Cortese, R. (1979). *J. Mol. Biol.* **129**, 169–172.
Smith, A. J. H. (1979). *Nucl. Acids Res.* **6**, 831–848.
Staden, R. (1977). *Nucl. Acids Res.* **4**, 4037–4051.
Staden, R. (1980). *Nucl. Acids Res.* **8**, 817–824.