Volume 11 Number 2 1983 Nucleic Acids Research

# Kilo-sequencing: an ordered strategy for rapid DNA sequence data acquisition

Wayne M. Barnes and Michael Bevan

Department of Biological Chemistry, Washington University School of Medicine, Saint Louis, MO 63110, USA

Received 29 September 1982; Revised and Accepted 10 December 1982

## ABSTRACT

A strategy for rapid DNA sequence acquisition in an ordered, nonrandom manner, while retaining all of the conveniences of the dideoxy method with M13 transducing phage DNA template, is described. Target DNA 3 to 14 kb in size can be stably carried by our M13 vectors. Suitable targets are stretches of DNA which lack an enzyme recognition site which is unique on our cloning vectors and adjacent to the sequencing primer; current sites that are so useful when lacking are Pst, Xba, HindIII, BglII, EcoRI. By an in vitro procedure, we cut RF DNA once randomly and once specifically, to create thousands of deletions which start at the unique restriction site adjacent to the dideoxy sequencing primer and extend various distances across the target DNA. Phage carrying a desired size of deletions, whose DNA as template will give rise to DNA sequence data in a desired location along the target DNA, may be purified by electophoresis alive on agarose gels. Phage running in the same location on the agarose gel thus conveniently give rise to nucleotide sequence data from the same kilobase of target DNA.

## INTRODUCTION

Gel sequencing methods for DNA are so rapid that a random approach has been used for several large sequencing projects [1,2] The most rapid random approach is that of Sanger and his collegues,[2] who sequence target DNA carried by M13, using the dideoxy method.[3,4] In this approach, randomly cleaved target DNA is molecularly cloned adjacent to a primer homology that can be conveniently used to acquire some 300 nucleotides of sequence data per sequencing experiment, several experiments per day. However, although some 1000 nucleotides of data can be acquired per day, this data is in disconnected blocks scattered randomly over both strands of the target DNA. A computer is then utilized to aid in the identification and ordering of each block ot sequence data [5-7]. Perhaps eighty percent of a large sequencing project may be determined fairly rapidly with this approach, before decreasing, repetitive returns set in.

We here describe a system for acquisition of data at the same peak rate (1000 nucleotides per man day), after initial engineering of the DNA so that





0305-1048/83/1102-0349$2.00/0

each experiment spans a predetermined region of the target sequence. Thus large sequencing projects can be attacked in a conveniently organized manner. Alternatively, small regions within large recloned restriction fragments can be accessed with maximum convenience for dideoxy sequencing (or other types of experiments which utilize deletions) without the need for more than minimal advance knowledge of the restriction site content of the DNA region.

We have applied this sequencing strategy to the T-DNA of the Ti plasmid T37 of Agrobacterium tumefaciens, starting from the right border, to which region map the nopaline synthase gene and some genes involved in plant cell transformation.

## METHODS

### Chemicals and Label

Deoxynucleoside triphosphates were obtained from Boehringer-Mannheim; dideoxy nucleoside triphosphates were supplied by PL Biochemicals. $\alpha$-$^{32}$P-dATP was supplied by Amersham or New England Nuclear at a specific activity of 400 Ci/mmole and diluted to 200 Ci/mmole at time of use. Ultra pure urea was from Schwarz-Mann, acrylamide from BDH. Other chemicals were reagent grade.

### Enzymes

Restriction enzyme FnuDII was a gift from A.Lui and M.Smith. Other restriction enzymes and Klenow enzyme (large fragment of DNA polymerase I, lot 5 or 9)[8,9] were from New England Biolabs. T4 DNA ligase was a gift from M. Bittner. Nuclease Bal 31[10] was supplied by New England Biolabs or New England Nuclear.

### Buffers and media

DNA buffer was 10 mM Tris-HCl pH7.9, 10 mM NaCl, 0.1 mM EDTA. Agarose gel electrophoresis buffer stock was 20XGGB = 0.8 M Tris-acetate pH 8.3, 0.4 M sodium acetate, 4 mM EDTA; it was diluted to 4XGGB for all gels except as indicated for the live phage gel. Rich medium for plates was NYE (14 g agar, 10 g NZ-amine A, 5 g yeast extract, 2.5 g NaCl per liter) and for liquid culture was 2XNY (25 g NZ-amine A, 10 g yeast extract, 1.5 g NaCl per liter); minimal medium salts were M9.[11] Reaction buffer NaTMS was 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 10 mM mercaptoethanol. Agar medium for plating M13 from the mWB series to form blue plaques contained per liter 12 g agar, 10 g NZ-amineA, 2.5 g NaCl, 40 mg XG (4-chloro-5-bromo-3-indolyl $\beta$-D-galactoside). When appropriate, antibiotics were added at the following concentrations: ampicillin, 100 ug/ml; kanamycin, 20 ug/ml; tetracycline, 20 ug/ml; streptomycin, 20 ug/ml.

Construction of WB373 ($tra^-$, $lac^+$, $M13^s$ host)

Strain WB368 was transduced to kanamycin and ampicillin resistance and $lacZ^-$ by infection with P1CMc1r100 grown on strain WB369, and a transductant was purified at 42° on lac-MacKonkey agar containing kanamycin (20ug/ml) and ampicillin (100ug/ml) and named WB370.

WB370 was crossed with MM294 and exconjugates were selected on minimal agar (M9 salts, glucose, $B_1$) containing kanamycin and ampicillin. The idea here was to select a transposition into the F plasmid that occurred in the minutes immediately before the mating. Some 500 to 1000 exconjugants were obtained per ml of mating mixture at $OD_{550}$ = 1. This was some 100 to 1000 fold fewer exconjugants than were obtained from a simultaneous control mating of WB369 x MM294. To screen for inactivation of the ability to transfer DNA ($Tra^-$), exconjugants from WB370 x MM294 were purified on antibiotic-containing lac-MacKonkey agar and pure, $Lac^+$ colonies were screened for potency by spot-mating with WB367 on LB agar containing streptomycin and kanamycin. At the same time, these exconjugants were tested for sensitivity to infection by M13 by smearing a small amount of each strain onto agar containing tetracycline and previously seeded with $10^{10}$ M13Ho168.3tet1. $M13^s$ clones were evident as $tet^R$ smears after 12 hours at 37°. WB373 is a $Tra^-$, $M13^s$ example from this screen. WB374 is $Tra^+$ $M13^s$.

Construction of mWB23 vector

Passenger DNA: 1.5 ug of mWJ22 RF DNA (a gift from R. Wu) were digested to completion with restriction enzyme FnuDII in NaTMS buffer. The reaction was then supplemented to 12 mM $CaCl_2$ and 200 mM NaCl, and 0.7 units of Bal 31 nuclease were then added and digestion allowed to proceed for one minute at 22°, after which the reaction was terminated with excess EDTA. Agarose gel analysis indicated that some 150 base pairs had been removed from each FnuDII DNA fragment. The DNA was heat-treated (70°, 10 minutes) and precipitated with ethanol. Neither of these last two treatments is certain to have inactivated the Bal 31 nuclease (see below).

Vector DNA: Wild-type M13 RF DNA was digested to completion with Sau96 restriction enzyme and then treated with DNA polymerase large fragment and all 4 dNTP's to fill in each end to the blunt form.

Ligation: 1 ug each of the above passenger DNA and vector DNA were ligated with T4 DNA ligase in a volume of 20 ul for 24 hours at 22°.

E.coli strain 71.18 was transformed with a portion of the ligation reaction above, and plated to detect plaques on agar supplemented with 40 ug/ml XG indicator and 1 uM IPTG. M13 carrying the N-terminal portion of the

Table 1. Bacteriophage strains

| Strain | Description | Source.Reference |
|---|---|---|
| P1CMc1r100 | | J. Davies[12] |
| M13 | wild-type | J. Kabori[13,14] |
| M13Ho168.3tet1 | BglII-fragment (tet$^r$ gene) of Tn10 carried at the edge of the histidine operon DNA carried by M13Ho168.3, a deletion (hsi603) of HindII fragments R7, R5, R3 of M13Ho168. | references 15,16,17 |
| mWJ22 | Carries the lac po HindII fragment inserted at nucleotide 5869 of M13. The early part of lacZ on this phage has a HindIII site bounded on each side by an EcoRI site. | reference 18 |
| mWB23 | 465 base pairs of the lac DNA of mWJ22 inserted at the Sau96 site (position 5725) of M13$^+$. | |
| mWB2341 | First EcoRI site of mWB23 eliminated. | |
| mWB2342 | Second EcoRI site of mWB23 eliminated. | |
| mWB2344 also known as mWB2348 | Polylinker HindIII, XbaI, BglII, PstI, EcoRI DNA from plasmid piVX added between EcoRI and HindIII sites of mWB2342. | |

Table 2. Bacterial strains

| Strain | Description | Source.Reference |
|---|---|---|
| WB351 | /\his-gnd Sm$^r$/F$^+$::Tn3 | reference 17 |
| 71.18 | /\lacpro/F'lac /\M15 i$^q$ pro | P.H.Hofschneider[19] |
| WB367[DB1512] | Sm$^r$ sup2 /\lacpro trp$^-$ | D.E.Berg |
| WB368[DB1683] | trp$^-$ /pOX38=F/\IS | D.E.Berg[20] |
| WB369 [DB129-234] | trp$^-$ /\lac-pro/ F'lacpro lacZ::Tn5-DR2 | D.E.Berg[21] |
| WB370 | lacZ::Tn5-DR2 trp$^-$/pOX38=F/\IS | |
| MM294 | F$^-$ endoI$^-$ Sm$^s$ supE 44 hsd R17 B$_1^-$ | M.Innis[22,23] |
| WB373 (tra$^+$version = WB374) | pOX38 tra::Tn5-DR2/ hsd R17 supE 44 B$_1^-$ Sm$^s$ endoI M13$^s$ | This is our M13$^s$ tra$^-$ lac$^+$ host. Its F plasmid is pWB373. |
| W3110pVXp3 | W3110 r$_c$m$_k$/plasmid piVX11 (polylinker, supF) + plasmid p3 (tet amber, amp amber) | B.Seed[24] |
| pBR325-Hind23 | Plasmid pBR325 carrying HindIII fragment 23 of Ti plasmid pTiT37 of Agrobacterium tumefaciens | M.-D. Chilton[25] |

β-galactosidase gene form blue plaques in this situation, by alpha complementation of deletion M15 in the host bacteria.[19] Several dozen blue plaques of various sizes, color density and genetic stability were picked and purified three times. One of these, mWB23, which appeared to be genetically

```
     5710      5720       .start lac DNA     5750      5760      5770      5780      5790      5800
ATTTGGGTGATGGTTCACGTAGTGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGG
                  end i gene.

     5810      5820      5830      5840      5850      5860      5870      5880      5890      5900
CTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGAATTCGCGCGACCGGATCCGGGCAACGTTGT
      <---------------lac operator                fMet Z gene...          BamHl**
        probe primer                                        EcoRI

This is the linker region for mWB2344; see below for 2341 and 2342.
     5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
TGCCATTGCTGCAGGCGCAGAACTGGTAGGTATGGAAGATCTCTAGAAGCTTGTGGAATTAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAA
       PstI                          XbaI HindIII                                  <---primer-----
                                  BglII                                            (N.E.Biolabs)

     6010      6020      6030      6040      6050      6060      6070      6080      6090      6100
ACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCCTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT

     6110      6120      6130      6140      6150      6160      6170      6180      6190      6200
GCGTAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTC

                                                lac DNA ends.*
     6210      6220      6230      6240      6250      6260      6270      6280      6290      6300
GTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA

     6310      6320      6330      6340      6350      6360      6370      6380...  total = ...6940
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTAT...     ...TTGGATGTT
F <------------------ F
       ori+ primer

.....

mWB23          5850      5860      5870      5880      5890      5900      5910      5920      5930
cloning region: AAACAGCTATGACCATGATTACGAATTCCACAAGCTTGTGGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACC
                      fMet Z gene... EcoRI    HindIII  EcoRI                  <---primer-----

mWB2341        5850      5860      5870      5880      5890      5900      5910      5920      5930
cloning region: AAACAGCTATGACCATGATTACGAATTAATTCCACAAGCTTGTGGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA
                      fMet Z gene...              HindIII  EcoRI                  <---primer-----

mWB2342        5850      5860      5870      5880      5890      5900      5910      5920      5930
cloning region: AAACAGCTATGACCATGATTACGAATTCCACAAGCTTGTGGAATTAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA
                      fMet Z gene... EcoRI    HindIII                          <---primer-----

       Total length of mWB2341 and mWB2342 is 6867.
.....
```

Figure 1. DNA sequence surrounding the cloning region of our M13 vectors mWB23, mWB2341, mWB2342 and mWB2344. Sequence numbering before the lac insert is that of the wild-type M13 sequence(14). The primer "N.E.Biolabs" is the one used for dideoxy sequencing of target DNA. The "ori$^+$ primer", a gift from Monsanto, is an M13 homology used in emergencies to sequence constructs which may have deleted the N.E.Biolabs primer lac homology. The indicated "probe primer" is useful for generating probe for hybridization screening (N.E.Biolabs).
* 8 nucleotides of the M13 origin of minus strand replication (ori$^-$) have been inadvertantly deleted at the 3′ boundary of the lac insert. Others have shown that this DNA is not essential to M13 growth (48).
** This BamHl site is not unique on the genome, as the M13$^+$ Bam site at 2220 is still there. The BglII site of mWB2344 is recommended for GATC ends.

stable, formed medium sized, medium blue plaques, and which was found by the orientation trick (using mWJ22 DNA as probe) to carry the anti-sense strand of the *lac* DNA (the *cal* strand), was chosen as the basis for further vector development. The DNA sequence of the *lac* cloning region of mWB23 is shown in figure 1. Note that 8 base pairs is missing from the M13 *ori*$^-$ DNA (minus strand origin of replication) adjacent to the Sau96 site on one side. This deletion may have been caused by incomplete inactivation of the Bal 31 nuclease.

*Construction of mWB2341 and mWB2342*

Each of the EcoRI sites of mWB23 were eliminated, one at a time, by means of partial digestion with EcoRI, filling in the ends with Klenow enzyme, religation, and transformation of *E.coli* 71.18. The DNA sequence of the *lac* region of these two phages is indicated in figure 1.

*Construction of mWB2344*

Additional cloning sites PstI, BglII, and XbaI were added to mWB2342 by incorporating the polylinker sequence of plasmid piVX.[24] The entirety of this plasmid was cloned *via* its single HindIII site into the HindIII site of mWB2342, and all of the smaller EcoRI fragments were eliminated by treatment of RF DNA with EcoRI and recircularization of the largest EcoRI fragment. The confirmed DNA sequence of the cloning region of mWB2344 is indicated in figure 1.

*Orientation trick*

The relative or absolute orientation of passenger DNA in M13 constructs was determined by annealling together the virion DNA from two clones at a time,[26] followed by agarose gel analysis: To 0.1 ml each of 2 infected-culture supernatants was added 50 ul of 20xGGB electrophoresis salts and 30 ul of toothpick assay lysis mix[27] (50% glycerol, 5% SDS, 0.1M EDTA, 0.1% bromphenol blue), and the mixture was annealled at 67°C for 1 to 16 hours.

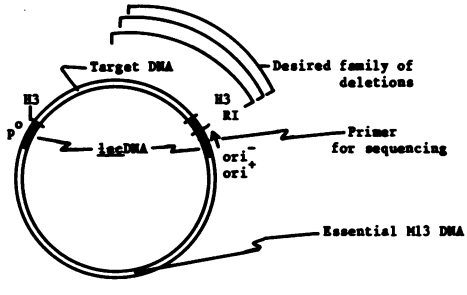


*Figure 2.* Starting material for the kilo-deletion procedure. A target HindIII fragment has been recloned so that a unique EcoRI site may anchor the deletions. The constructs carrying the two orientations of the target Ti DNA HindIII fragment 23 were named mTi23.11 and mTi23.2. The vector was mWB2341.

Each phage was analyzed identically but separately as well. Agarose gel analysis clearly showed two slower-moving bands when the two DNA's carried opposite homologous strands. If purified DNA from a well-characterized M13 clone was available, 1 ug of this probe DNA was added instead of phage suspension.

Cloning of target DNA

MWB2341 vector RF DNA was treated with HindIII and calf intestine alkaline phosphatase.[28] HindIII fragment 23 (3.4 kb) of Ti plasmid DNA was recloned from pBR322 into the HindIII site of mWB2341. Clones having the expected genome size (10 kb), according to the toothpick assay of phage-infected bacterial smears, were put into two orientation classes by the orientation test described above. Clone mTi23.11 carried fragment HindIII 23 in the orientation with the right border of T-DNA toward the side with the

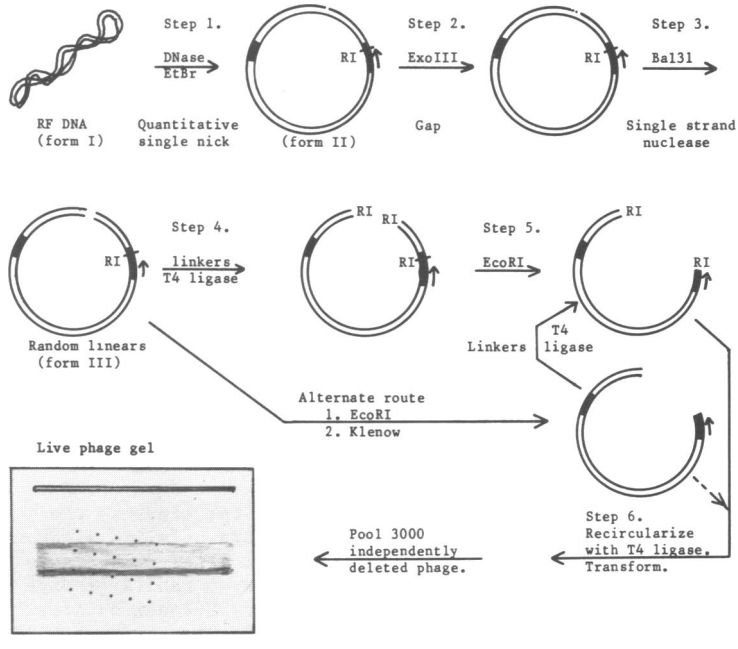


Figure 3. The steps in the kilo-deletion procedure. The RF DNA described in figure 2 was subjected to the enzymatic treatments shown. A total of three routes to the final deletions are indicated. Two of the routes end up with linkers at the point of the deletion. The route indicated by the dotted line does not utilize linkers and may sometimes be preferable, depending on the quality of available linkers, T4 ligase, and skill. Monitoring of each step is recommended as demonstrated in figure 4. An example of a live phage gel is shown in figure 5.

sequencing primer. Clone mTi23.2 carried the opposite orientation, with the HindIII site internal to T-DNA adjacent to the sequencing primer. RF DNA was prepared from one liter of infected cells by the cleared lysate method,[29] followed by two bandings in CsCl/ethidium bromide.

Creation of deletions in vitro -- The Kilo deletion procedure

Step 1. Nick once with DNAse.[30] The reaction buffer for the nicking reaction contained 125 mM NaCl, 20 mM $MgCl_2$, 4 mM Tris-HCl pH 7.9, 60 ug/ml BSA, and 500 ug/ml ethidium bromide. 5 ug RF DNA were treated with 50 ng pancreatic DNAse I in a volume of 100 ul at 25° for 1 hour. Agarose gel analysis of a 0.2 ug aliquot (see figure 4, channel 2) showed 100% disappearance of the supercoiled DNA (RFI), 90-100% conversion to nicked circles (RFII), and 0-10% linears (RFIII). The DNA was extracted twice with

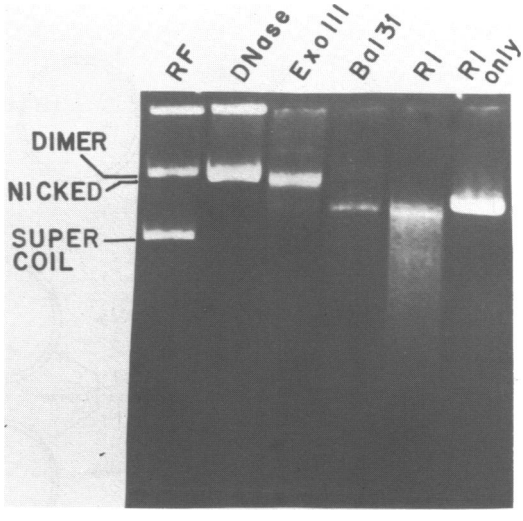


Figure 4. 1% agarose gel analysis of some of the steps in the kilo-deletion procedure. Channel 1: Untreated RF DNA of mTi23.2. Channel 2: After DNAse + ethidium bromide nicking. Channel 3: After exonuclease III treatment of the randomly nicked DNA. Some batches of exoIII produce some unexpected (but harmless) linears in this reaction. Channel 4: After Bal 31 single-strand nuclease treatment of the gapped DNA. The apparent 50-70% loss of material is typical. Channel 5: After EcoRI treatment of the randomly linearized DNA in channel 4. This reproduction might not show it, but superimposed upon the expected smooth smear of DNA fragments there is a more intense band slightly smaller than linears. This may be due to preferential (non-random) nicking by DNAse at a position near the EcoRI site in the M13 DNA. A large block of high AT near the ori$^+$ may be this site. We have not further investigated this phenomenon, partly since these molecules are probably not viable. Channel 6: Original RF DNA treated with EcoRI to show the gel position for linears.

phenol, twice with chloroform, and precipitated with ethanol.

Step 2. Widen nick to small gap with exonuclease III. The nicked DNA from step 1 was resuspended in 50 ul exoIII buffer (66 mM Tris-HCl pH 7.9, 0.6 mM $MgCl_2$, 0.1 mM mercaptoethanol) and treated with 20 units of exonuclease III (New England Biolabs) at 25° for 10 minutes. The enzyme was inactivated by heat treatment at 70° for 10 minutes. Agarose gel analysis (figure 4, channel 3) showed no significant change from step 1.

Step 3. Cut across gap and further randomize ends with nuclease Bal 31. The entire exoIII reaction was diluted to 200 ul in Bal 31 buffer (final 0.6 M NaCl, 12 mM $CaCl_2$, 12 mM $MgCl_2$, 25 mM Tris-HCl pH 7.9, 1 mM EDTA). 0.06 units of Bal 31 nuclease (New England Biolabs), an amount previously determined to remove about 3 base pairs per minute from the ends of fragments of 5 ug of a lambda RI digest, were then added. After 5′ at 25°, 1/2 of the reaction was stopped by adding it to excess EDTA. This treatment was sufficient to convert all of the gapped circles to linears (see figure 4, channel 4). 0.12 units of Bal31 were then added to the remaining 1/2 of the reaction mixture and incubation was continued for another 5 min. 0.2 ug of each Bal 31 time point were analyzed on an agarose gel, and then both time points were pooled and deproteinized as for step 1.

Step 3a. Polish ends with DNA polymerase large fragment to be sure they are blunt. DNA from step 3 was reacted in 50 ul NaTMS buffer with 100 uM all four dNTP′s and 1.5 units of DNA polymerase I large fragment (New England Biolabs) for 15 minutes at 37°. Deproteinization was applied as for step 1.

Step 4. Linkers (5′ CGGAATTCCG 3′) were phosphorylated by treatment with T4 polynucleotide kinase and $\gamma$-$^{32}$P-rATP. They were then used directly after heat-inactivation of the kinase. In a reaction volume of 20 ul, one half microgram of phosphorylated EcoRI linkers were ligated to 4 ug of randomly linearized RF DNA molecules using T4 DNA ligase under blunt-end ligase conditions (30 mM Tris-HCl pH7.9, 10 mM DTT, 4 mM $MgCl_2$, 1 mM EDTA, 1 mM rATP, 1 mM hexamminecobalt chloride[31]). Incubation was for 16 hours at 4° C. Deproteinization was then applied as for step 1. ATP and excess linker were partially removed by gel filtration on a 1 ml Biogel A-15m column[2].

Step 5. Digest with the unique enzyme. The DNA was digested with a 10-fold excess of EcoRI restriction enzyme. It is this digestion which created the desired deletions.

Step 6. Recircularization. 1 ug of EcoRI-digested RF DNA which survived steps 1-5 was ligated under sticky-end ligation conditions with T4 ligase in a

reaction volume of 100 ul. The reaction was incubated at 4° for 16 hours.

Transformation of competent E.coli WB373 was as described.[32] Transformed cells and fresh lawn cells were plated by the soft agar method on agar containing XG at 40 ug/ml. Desired phage formed blue plaques.

Live phage gel

Phage plaques from the transformation (ca. 3000) were suspended together in 5 ml M13 buffer, clarified by centrifugation, heated at 70° for 10 minutes, and this phage deletion pool was then stored at 4°. A gel sample was assembled at 60° by mixing 0.3 ml phage pool, 0.2 ml blue2 (50% glycerol, 0.1 M EDTA, 0.1% bromphenol blue, 0.1% xylene cyanol), and 0.3 ml 0.7% molten agarose. This gel sample was then loaded into the bottom 2/3 of a wide slot of a 20cm x 20cm x 2.5mm horizontal 0.7% agarose (high-melting-temperature) gel. The phage sample was then overlayed with 0.7% agarose and the horizontal gel covered with Saran wrap. Reservoir geometry and wick contact with the electrode reservoirs was as described,[33] except that the gel was covered with Saran wrap. The electrophoresis buffer was 4XGGB. 100 mA of current (3 V/cm) were applied for 42 hours with no overt cooling. Later experiments utilized 3XGGB for 20 hours. After electrophoresis, the gel was sampled at 1.8 mm intervals with a "polypoker" (available from Swell Gels, St. Louis), a 5 x 8 array of blunt syringe needles at 9 mm spacing, which corresponds to the spacing of the wells of a standard microtiter plate, into which the phage and agarose picked up by the needles were rinsed with 0.2 ml M13 buffer. After sampling the gel, the phage left in the gel were deproteinized by soaking in 0.1% SDS for 30 minutes, followed by 4 changes of 0.5xGGB to remove the SDS. Later experiments used 0.1 N NaOH for 30 min to deproteinize, followed by neutralization with 4XGGB. The gel was then stained in 0.5xGGB containing 0.5 ug/ml ethidium bromide and photographed under shortwave UV illumination as described.[27] Only after inspection of this photograph (figure 5) was it apparent which gel samples contained desirably deleted phage.

DNA sequencing

Isolation of template DNA was nearly as described by Sanger et al.,[2] except that infected cultures were incubated overnight, and the phage from 3 ml of culture supernatant were extracted to yield about 19 ug DNA.

The dideoxy chain-termination method of Sanger et al.[3] was employed, with the addition of a ddGTP/dITP channel (ddGTP:dITP::2.5:1) to aid in the resolution of gel compressions.[31] Primer DNA was a 15-mer lac homology of the sequence 5'TCCCAGTCACGACGT 3', supplied by New England Biolabs.

DNA sequence data was reduced from autoradiographs with the aid of a

semi-automated gel reader[34] (available from Swell Gels). We describe many of these methods in detailed recipe format elsewhere.[35]

STRATEGY

Recloning the target DNA so that a unique restriction site is on one side, adjacent to the sequencing primer.

Our strategy involves the initial cloning of large (3 to 14 kb) regions of target DNA into M13 cloning vectors which can stably carry such large amounts of passenger DNA. Care is taken that the resulting construct have a unique restriction site at only one edge of the target DNA, adjacent to the sequencing primer homology, as shown in figure 2 for our starting material.

This step has basically two approaches. One approach is to clone a target DNA fragment into a site or combination of sites which is on the other side (vis-a-vis the sequencing primer) from a unique site in the linker region of the vector. The unique site should be one which does not appear in the target DNA. For instance, if the target DNA lacks an XbaI site, one clones the target DNA into a site on the other side of the XbaI site, such as into the PstI, BglII, or EcoRI site of mWB2344. The ends of the target DNA in this approach may be irrelevant; they can be a partial Sau 3A digest for insertion into the BglII site of mWB2344, or they can even be a random DNAse cleavage product. Both orientations will usually be obtainable and desirable, so that the entire deletion and sequencing strategy can derive sequence data and biological expression data on both strands, if necessary. The two possible orientations of the target DNA can be conveniently distinguished by annealling together phage suspensions, followed by agarose gel analysis, as described under 'orientation trick' in Methods. Another approach is to take advantage of restriction sites in or near the target DNA sequence that have sticky ends that match those in the mWB2344 polylinker region. In this case the orientation of the target DNA can be forced. Only one sticky end need be used to force the orientation, if the other end of both the passenger DNA and target DNA is first polished to a blunt end by treatment with the large fragment of DNA polymerase. If a sticky end used for this cloning approach is on the primer side of the passenger, this restriction site can function as the near-side deletion boundary in the kilo-sequencing strategy.

Selection for deletions in the desired direction. Note that cleavages in the M13 portion of the RF DNA will result in non-viable molecules.

Add linkers to the random ends. The endpoints of the deletion may be joined without linker if the Klenow enzyme polishing step is moved to after

digestion with the unique enzyme, and if the final recircularization is carried out under blunt ligation conditions (higher DNA and enzyme concentration). On a statistical basis, 25% of such blunt ligations may be expected to regenerate[37] the restriction site anyway, for the enzymes EcoRI, HindIII, XbaI and BglII, but not PstI.

The use of linkers has several advantages over the alternate, perhaps more reliable, procedure of blunt-end ligation at the recircularization step: First, the recircularization ligation is easier to carry off efficiently, so that more deleted phages may result per microgram of manipulated DNA. Second, the ligation can be performed under more dilute conditions than are required for blunt-end ligation[38] so that the danger of undesired intermolecular ligation events is reduced. In corollary, any undesired intermolecular ligation events that do result can be recognized, if necessary, since they will create more than one restriction site for the unique enzyme being used. Finally, and most importantly, the linkered restriction site at the boundary of the deletion may be useful in other biological experiments.

## RESULTS

### M13 vectors constructed for use with the method

M13 vectors were constructed to have the following features, many of which are not available for other vector series:

1) They carry foreign DNA adjacent to the M13 origin of minus strand replication. Empirically (see Discussion) M13 vectors that carry DNA at this location are much less subject to deletion, that is, foreign DNA is carried more stably by these vectors than by, for instance, the mp series of Messing et al.[19,39] which carry foreign DNA in a HaeIII site on the other side of the intergenic region.

2) They have *lac* homology from the early part of the β-galactosidase gene of *E. coli*, in order that they may use the primer made commercially available from several sources for use with the mp series.

3) Adjacent to the commercially available primer is a polylinker containing one each of several restriction sites, including, to date, EcoRI, PstI, BglII, XbaI, and HindIII.

4) They carry the *lac* operator, which, when it binds up *lac* repressor, leads to induction of the chromosomal wild-type *lac* operon[37] in the host *E.coli* WB373, rendering all the plaques blue on XG indicator agar. This has two useful functions. (a) It makes the plaques easier to see, a valuable feature for constructs carrying large amounts of foreign DNA, which form small

plaques and (b) it signals when deletions constructed for the sequencing strategy have gone too far, since deletions which have cut as far as the *lac* operator, i.e. clear across the target DNA, cause the plaques to be colorless.

*Host E.coli* strain WB373

A $Tra^-$, $M13^s$, $Lac^+$ *E.coli* host strain was constructed as described in Methods by transposing Tn5-DR2 into a *tra* gene of the small F plasmid pOX38.[20] The transposon Tn5-DR2 was chosen for this purpose because it is known to excise at a low frequency ($3 \times 10^{-9}$) some 300 times lower than for wild-type Tn5.[21] The $Tra^-$ defect in the F plasmid of WB373 was found to be complete, since, with WB367 as the female, less than one exconjugant was found to arise from a large scale (1 ml + 1 ml of late log phase cultures) mating under conditions that allowed the $Tra^+$ strain WB374 to produce $3 \times 10^9$ exconjugants. This mating selected for transfer of $kan^r$ and $amp^r$ of Tn5-DR2 carried by the F plasmids pWB73 or pWB74. Thus, under antibiotic selection, WB373 was found to be adequately transfer-deficient for genetic engineering

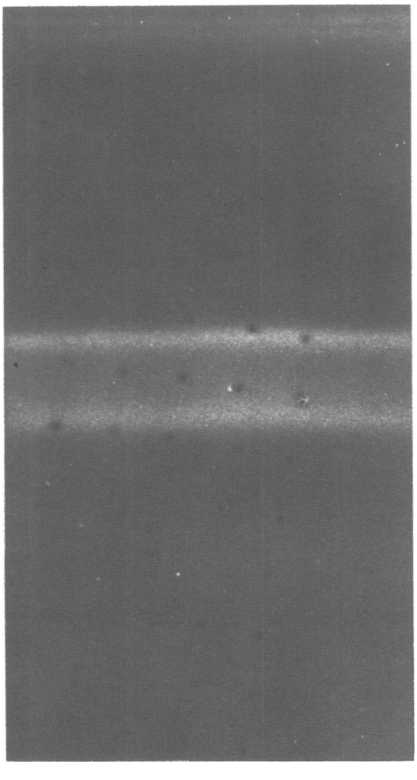

Figure 5. Live phage gel separation of deletions of mTiz3.11. For scale, the pins of the polypoker are at 9 mm intervals. Only phage adhering to pins 6-10 (the second row) were studied further.

purposes that require this feature of biological containment. The transfer of possible excision products from WB373 grown in antibiotic-free medium was not tested.

Deletion spectrum. The overall performance of the method for generating deletions may initially be judged from the live phage gel in figure 5. The live phage gel displayed a prominent band at the position of undeleted phage in the particular experiment shown. These undeleted phage would result from inefficiency at nearly any of the enzymatic steps of the deletion procedure, although in the experiment shown, the only step which did not go to completion, according to agarose gel analyses throughout the procedure (results not shown) was the last EcoRI digestion. We believe (but have not proved) that the heavy band at the fully deleted position results from the previously observed[41] unusual sensitivity of the lac operator to single strand nucleases. These fully deleted phage formed colorless plaques, and they could therefore be easily ignored. The phage between the two heavy bands were the desired deletions, and they were sampled on pins 5-10 of the polypoker, as shown by the holes in the live phage gel (figure 5).

As a routine, we recommend that 10 purified clones of phage from each point of the live phage gel be sized more accurately with the toothpick assay.[27] Several phage from each pin in the experiment in figure 5 are so analyzed in figure 6.

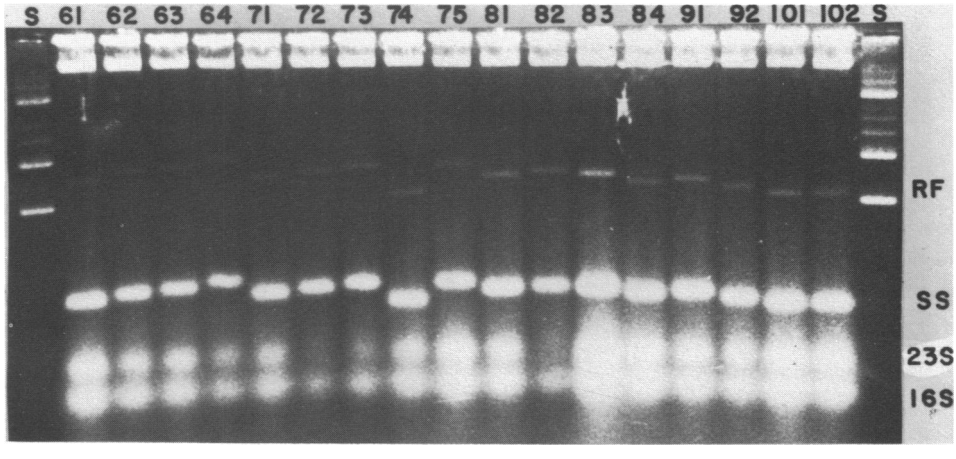


Figure 6. Phage from the live phage gel in figure 5 were analyzed for genome size by the toothpick assay(27). Deletions numbered in the sixties are from pin number 6, etc. The outside lanes are standard RF DNA's, the smallest two of which are 6867 and 9738 b.p. in length. As an example, phage number 73 is also appears in figure 7.

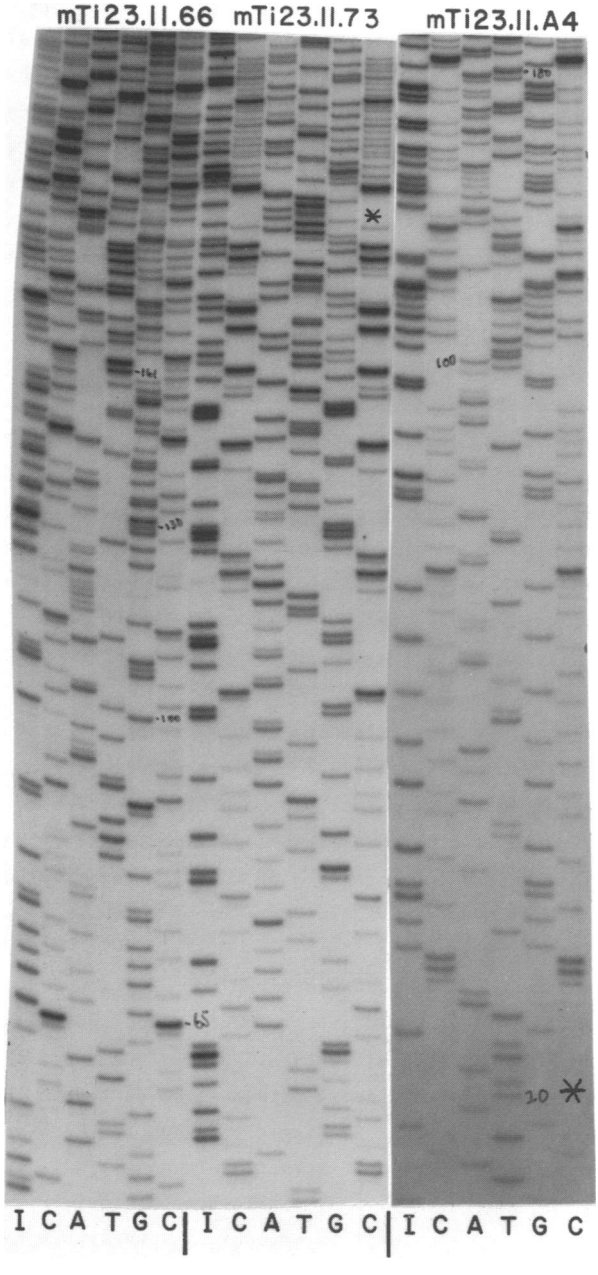


Figure 7. Example dideoxy sequencing gels resulting from the kilo-sequencing strategy. The phage used for the experiment on the left, mTi23.11.66, is actually not deleted. The data from deletions 73 and A4 overlap in the regions indicated by the asterisk (*).

The final test of the method was DNA sequence determination using as templates the phage which gave a promising genome size result in the toothpick assay. In the event, phage with marginally different genome sizes in the toothpick assay did indeed give rise to adjacent, overlapping sequence data. A couple of sequencing experiments from two deletions which gave overlapping data are shown in figure 7.

We present and discuss the sequence itself in the adjacent paper.[42] It covers 3000 nucleotides from the right portion of the T-DNA of the Ti plasmid T37 of Agrobacterium tumefaciens, including the right border of T-DNA and the entire nopaline synthase gene nop.

## DISCUSSION

We have described and demonstrated a strategy for conveniently creating thousands of deletions which are anchored at one edge of a target sequence and extend various distances across that sequence. When, as in the demonstrated case, the anchor point is a primer homology carried on M13 phage, then the target DNA sequence may be determined by the dideoxy method in an organized manner. Once, or even before, the DNA sequence is known, other biological experiments may utilize the deletions to locate genes or DNA recognition sites.

Not only is this strategy valuable for organizing the study of a large sequence (about 10 kb); it is also a convenient way to access small (1 kb) regions within a larger sequence that has been cloned onto our M13 vectors: The deletions essentially move the sequencing primer over to the gene of interest, avoiding the random sequencing of uninteresting regions, or avoiding more detailed subcloning or restriction site mapping.

### Advantages of this system.

Organized strategies have been described[43,44] for sequencing DNA by the chemical method, using plasmid vector. Our method adds organization to the relative conveniences available with M13 vector and the dideoxy method. These conveniences are 1) The dideoxy method involves little labor and no noxious or dangerous chemicals. 2) Single-stranded template for the dideoxy method is easily prepared by phenol extraction from just a few milliliters of M13-infected culture supernatant. 3) Single-stranded primer which can be used for each sequencing experiment is economically available commercially.

While this paper was in preparation there appeared two other similarly organized sequencing strategies which utilize M13 and the dideoxy method.[45,46] We wish to enumerate the superior advantages of our system.

Stability of target sequences up to 14 kb (at least). The vectors mWB2341,42 and 44 constructed for the system described here can carry large target DNA sequences, at least up to 14 kb, with little or at least greatly reduced propensity to form unwanted deletions. We have cloned the entire histidine operon from Salmonella typhimurium, a 10 kb PstI fragment, into the PstI site of mWB2344.[47] We have also cloned a 12 kb BamHl fragment (Bam6 of T-DNA) into the BglII site.[48] We are able to grow mg amounts of DNA from these constructs without noticeable ($\leq$1%) deletions, according to agarose gel analysis.

Our choice of genomic location for cloning on M13 stemmed from previous empirical evidence. We had found that histidine DNA was stably carried at the HaeIII site adjacent to the minus strand origin of replication.[16,26] In contrast, when we have tried to subclone smaller (3-4 kb) regions of the histidine operon into another vector, M13mp2, we found it difficult or impossible to grow quantitatve amounts of DNA without many (>50%) smaller-sized genomes evident in the preparation. We therefore sought to use the ori$^-$ genomic location, which has a unique Sau96 site in the wild-type M13, for the current series of vectors. We do not know why large target DNA is more stable in our vectors, except to note that our vectors carry foreign DNA on the other side of the origins of replicaton of M13 than do the mp series. This, or the fact that we may have inactivated the minus strand origin of replication by the inadvertant deletion of 8 base pairs,[49] may, through some unexplained effect on DNA replication, select against deleted molecules or prevent the deletions in the first place.

We wish to report that we have observed two types of foreign DNA that we cannot clone stably even at the ori$^-$ location of M13. One type of such DNA carries a strong promoter that drives transcription in the direction opposite to M13's own transcription.[17] In fact the phage mWB23, with its lac promoter, is a weak example of such a case. The presumably polar frameshifts introduced to create mWB2341 and mWB2342 resulted in faster-growing, more stable phage. The other DNA that we are unable to propagate stably in M13 is transposable element DNA such as Tn10 or IS1.

Direct size selection for phage carrying desired deletions. Our procedure contains no step where DNA must be purified and eluted from a gel. We do have a gel purification step for deletions of desired sizes, but it is an optional step which occurs at the end of the procedure, and it is carried out upon highly infectious phage particles, which need not be eluted with high efficiency, nor processed with any enzymes. This size purification is not

overly accurate, since only 50% of the phage from a given point on the gel are within 1 kb in genome size.

As a final, fine determination of the sizes of the deletions, we employ the toothpick assay which can be carried out accurately on infected smears of bacteria, without the need to genetically purify, grow up larger cultures, purify any DNA or carry out any exploratory sequencing reactions, such as T channels. In the toothpick assay, we base our size determination on the mobility of the RF DNA, rather than the single-stranded DNA, which, in accord with general experience for single-stranded RNA, does not allow accurate size determination based on mobility, unless a denaturing gel is used. We find that a sizing accuracy of 1% is possible, if RF DNA size standards are included within each gel sample.

Color screen against deletions which are too large. All of our phage form blue plaques if they carry the lac operator, which is located on the far side of the target DNA vis-a-vis the sequencing primer and the anchor point for the deletions. Although the live phage gel allows one to avoid most deletions which are too large, those that do come through this size purification are easily ignored because they are colorless.

Intramolecular ligation. In contrast to the procedure of Poncz et al,[45] and in common with the procedure of Hong,[46] we employ intramolecular ligation, having cut the RF DNA once randomly and once specifically. This allows efficient ligation. Our yield of independent deletions per microgram of RF DNA is two orders of magnitude higher than that obtained by Hong.[46] With further biological experiments in mind, we recommend that linkers be utilized when recircularizing to form the final constructs. This will allow later cision at the point of any interesting deletions that are sequenced.

ACKNOWLEDGEMENTS

We thank M.-D.Chilton for subcloned Ti-DNA, Paula Son and Elodee Tuley for technical assistance, G.Gallupi (Monsanto) and N.E.Biolabs for gifts of primer, and the NIH (grant no. GM24956) and the American Cancer Society (Faculty Research Award to W.B.) for support.

REFERENCES

1. Sutcliffe, J.G. (1979) Cold Spring Harbor Symposium 43, 77-90.
2. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J.Mol.Biol. 143, 161-178.
3. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Nat.Acad.Sci.USA, 74, 5463-5467.

4. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
5. Staden, R. (1979) Nucl. Acids Res. 6, 2601-2610.
6. Staden, R. (1980) Nucl. Acids Res. 8, 3673-3694.
7. Gingeras, T.R., Milazzo, T.J.P., Sciaky, D. and Roberts, R.J., (1979) Nucl. Acids Res. 7, 529-545.
8. Klenow, H., Overgaard-Hansen, K. and Patkar, S.A. (1971) Eur. J. Biochem. 22, 371-381.
9. Setlow, P., Brutlag, D. and Kornberg A. (1972) J.Biol.Chem 247, 224-231.
10. Legerski, R.J., Hodnett, J.L. and Gray, H.B. (1978) Nucleic Acids Research 5, 1445-1463.
11. Miller, J.H., Experiments in Molecular Genetics, Cold Sring Harbor Laboratory, Cold Springs Harbor, N.Y., 1972, pp. 431-433.
12. Mojica-A, T. (1975) Molec. Gen. Genet. 138, 113-126.
13. Hofschneider, P.H. (1963) Z.Naturforsch. 18b, 203-210.
14. Van Wezenbeck, P.M.G.F., Hulsoebos, T.J.M., and Schoenmakers, J.G.G. (1980) Gene 11, 129-148.
15. Jorgensen, R.A. and Reznikoff, W.S. (1979) J. Bacteriol. 138, 705-714.
16. Barnes, W.M. (1979) Gene 5, 127-139, and unpublished work.
17. Barnes, W.M., and Tuley, E., (1983) J.Mol.Biol. in press.
18. Rothstein, R.J., Lau, L.F., Bahl,C.P., Narang, S.A. and Wu, R., in Methods in Enzymology vol 68, ed. Wu, R. (1979) Academic Press,, 98-109.
19. Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) Proc.Nat.Acad.Sci.USA, 74, 3642-3646.
20. Guyer, M.S., Reed, R.R., Steitz, J.A., and Low, K.B. (1981) Cold Spring Harbor Symposium XLV, 135-140.
21. Egner, C. and Berg, D.E. (1981) Proc.Natl.Acad.Sci.U.S.A. 78, 459-463.
22. Durwald, H. and Hoffman-Berling, H. (1968) J.Mol.Biol. 34, 331-346.
23. Meselson, M., and Yuan, R. (1968) Nature 217, 1110-1118.
24. Seed, B., personal communication.
25. Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M. and Chilton, M.-D., (1982) Proc.Nat.Acad.Sci.USA 79 6322-6326.
26. Barnes, W.M. (1980) Genetic Engineering, vol 2, eds Setlow, J.K. and Hollaender, A. , Plenum, New York, 185-200.
27. Barnes, W.M. (1977) Science, 195, 393-394.
28. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J., and Goodman, H.M. (1977) Science 196, 1313-1316.
29. Katz, L, Kingsbury, D.T. and Helinski, D.R., J. Bacteriol. (1973) 114, 577-591.
30. Greenfield, L., Simpson, L. and Kaplan, D. (1975) Biochim. Biophys. Acta, 407, 365-375.
31. Barnes, W.M., unpublished.
32. Cohen, S.N., Chang, A.C.Y. and Hsu, C.L. (1972) Proc.Nat.Acad.Sci.USA 69, 2110-2114.
33. Shinnick, T.M., Lund, E., Smithies, O. and Blattner, F.R. (1975) Nucl. Acids Res. 2, 1911-1921.
34. Gingeras, T.R., Rice, T., and Roberts, R.J., (1982) Nucleic Acids Res. 10, 103-113.
35. Barnes, W.M., Bevan, M., and Son, P., Methods in Enzymology, in press. Nucleic Acids Res. 9, 3015-3027.
36. Barnes, W.M. (1978) J.Mol.Biol. 119, 83-89.
37. Backman, K., Ptashne, M. and Gilbert, W. (1976) Proc.Nat.Acad.Sci.USA 73, 4174-78.
38. Sugino, A., Goodman, H.M., Heyneker, H.L., Shine, J., Boyer, H.W., and Cozzarelli, N.R. (1977) J.Biol.Chem 252, 3987-3994.
39. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
40. Gronenborn, B. and Messing, J. (1978) Nature, 272, 375-376.

41. Chan, H.W. and Wells, R.D. (1977) Nature 252, 205-209.
42. Bevan, M., Barnes, W.M., and Chilton, M.-D., (1983) Nucleic Acids Res.
43. Frischauf, A.M., Garoff, H., and Lehrach, H. (1980) Nucleic Acids Res. 8, 5541-5549.
44. Hartley, J.L., Chen, K.K., and Donelson, J.E., (1982) Nucleic Acids Res. 10, 4009-4025.
45. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., and Surrey, S. (1982) Proc.Nat.Acad.Sci.USA, 79, 4298-4302.
46. Hong, G.F. (1982) J.Mol.Biol. 158, 539-549.
47. Barnes, W.M. and Son, P., unpublished data.
48. Vanderleyden, J., Bevan, M., Chilton, M.-D. and Barnes, W.M., work in progress.
49. Kim, M.H., Hines, J.C., and Ray, D.S. (1981) Proc.Nat.Acad.Sci.USA 78, 6784-6788.