Volume 8 Number 10 1980 Nucleic Acids Research

Nucleotide sequence of the E. coli gene coding for dihydrofolate reductase

Douglas R.Smith and Joseph M.Calvo

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

Received 12 February 1980

ABSTRACT

Plasmid pLC1437a contains DNA from Escherichia coli K12 including fol, the structural gene for dihydrofolate reductase. The fol gene was mapped on this plasmid relative to several restriction endonuclease cleavage sites. fol was also cloned from strain RSO and the nucleotide sequence for the entire fol gene and its flanking regions from this strain was determined. The amino acid sequence predicted from the nucleotide sequence differs in only a few respects from the reported amino acid sequence of dihydrofolate reductase from E. coli B. The major RNA transcripts initiated at the fol promotor in vivo are approximately 550 and 590 nucleotides long. In addition to these, several longer transcripts (up to 1400 nucleotides) are present in lesser amounts. A new procedure is described for 3' end labeling of DNA fragments having blunt ends using E. coli exonuclease III and avian myeloblastoma virus reverse transcriptase.

INTRODUCTION

Dihydrofolate reductase (DFRase, EC 1.5.1.3) catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate. A major function of this enzyme is to regenerate tetrahydrofolate from the dihydrofolate that is formed in the conversion of deoxyuridylic acid to deoxythymidylic acid (dTMP).

Drugs such as trimethoprim and methotrexate, which bind tightly to and inhibit DFRase, are useful clinically as antibacterial, antiprotozoal, immunosuppressant, and antineoplastic agents (1). Because these drugs have such widespread clinical significance, considerable work has been done on mechanisms by which cells become resistant to their action (2-5). Bacterial cells develop resistance to trimethoprim by acquiring a plasmid (6) or by mutation leading to an altered DFRase (7) or to increased synthesis of DFRase (7,8).

Our work is concerned with drug resistance brought about by mutation and, in particular, with aspects that relate to the control of synthesis of DFRase. A number of mutants of E. coli have been isolated that produce high levels of DFRase constitutively (7,8). The mutation in one of these, fol-60, is close-

ly linked to the structural gene for DFRase, fol, and is cis dominant (7). Furthermore, some strains carrying fol-60 are temperature sensitive for growth (7), and at least some revertants able to grow at the non-permissive temperature produce an altered DFRase (9, and our unpublished results). As a first step in understanding the nature of mutations such as fol-60, we identified a derivative of plasmid ColE1 that carries fol (6). In this paper, we report the nucleotide sequence of fol from the parent of the above mutants, strain RSO, and the sequence of nucleotides flanking fol. During the course of these studies, we developed some procedures that facilitate the sequencing of DNA.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

E. coli strains RSO (7), RS35 (7), JA200 (11), and M94 (6) have been described. Strain BL322, obtained from F.W. Studier, is a $nadB^+$, $purI^+$ transductant of strain PA3306 (10). Strains CV641 (M94/pLC1437a, constructed by transformation), CV628 (M94/pCV27), plus clones 04-46 and 14-37 of the Clarke-Carbon colony bank (11) were used as sources of plasmid DNA. The following selective media were used: ampicillin, 100 µg/ml in L-agar (12); tetracycline, 20 µg/ml in L-agar (12); trimethoprim, 2 µg/ml in M9-casamino acids (6); Colicin E1, 0.1 ml of a preparation (6) of titer 100 spread on the surface of an L-agar plate. Competent cells of strain M94 were prepared as described (6).

DNA Isolation

Plasmid DNA was prepared as described previously (6). Chromosomal DNA was prepared from a 50 ml L-broth culture of strain RSO (A_{550} = 1.0, Zeiss spectrophotometer) as follows. The cells were washed in M9 salts (without $MgCl_2$ and $CaCl_2$) (12), resuspended in 5 ml of the same solution and lysed by the addition of 10 ml of buffer containing 10 mM Tris-Cl, pH 8, 10 mM NaCl, 10 mM EDTA, and 1% sodium dodecylsulfate. The mixture was extracted twice with phenol saturated with TE buffer (10 mM Tris-Cl, pH 8, 1 mM EDTA), three times with diethyl ether, and the nucleic acid was precipitated with 0.1 volume of 3M sodium acetate and 2 volumes of 95% ethanol. This preparation contained a substantial amount of RNA in addition to chromosomal DNA.

Restriction Endonuclease Digestions

Restriction endonuclease digestions for analytical purposes, containing 1 µg of DNA and one quarter unit of enzyme (New England BioLabs) in 10 µl of

the buffer recommended by the manufacturer, were incubated 2-4 h at 37°C. Preparative scale digests containing buffer, 0.1-1 mg of DNA (at a concentration of 1 mg/ml), and 25 units of enzyme per mg of DNA, were incubated for 16-24 h at 37°C. One fifth volume of 50% glycerol containing 0.1% xylene cyanol and 0.1% bromophenol blue was added to samples prior to loading onto gels.

Contruction of Plasmids for Location of *fol* on Plasmid pLC1437a

Plasmids pCV42 and pCV39 (Table II) were constructed simultaneously as follows. Four µg of pBR322 DNA and 6 µg of pLC1437a DNA were digested with restriction endonucleases *Bam*HI and *Bgl*II. The digests were mixed together and heated for 10 min at 65°C. The sample was brought to 50 mM Tris-Cl, pH 8, 25 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, 100 µM ATP in a volume of 40 µl. One-tenth unit of T_4 DNA ligase (New England BioLabs) was added and the sample was incubated for 16 h at 12.5°C. $CaCl_2$ was added to a concentration of 100 mM and one-half of the mixture was used to transform 0.2 ml of competent *E. coli* strain M94 as previously described (6). The other plasmids in Table II were constructed in a similar fashion except that the restriction endonucleases indicated in Table II were used. Transformants were selected as follows: ampicillin resistance for clones containing pCV36, pCV37, pCV40, and pCV42; tetracycline resistance for clones with pCV43; colicin resistance for clones pCV38, pCV39, and pCV41. Clones were screened for the other drug marker on plasmid pBR322: tetracycline sensitivity for clones with pCV36, pCV37, and pCV42; tetracycline resistance for clones with pCV40; ampicillin sensitivity for clones with pCV43; ampicillin and tetracycline sensitivity for clones with pCV38, pCV39, and pCV41. To further characterize transformants, some of them were lysed by treatment with sodium dodecylsulfate and analyzed on agarose gels (13) to determine the approximate size of the plasmids carried by those clones.

Cloning of *fol* from Strain RSO

Twenty µg of a chromosomal DNA preparation from strain RSO was digested with restriction endonuclease *Bam*HI. This was mixed with 2 µg of *Bam*HI-digested pBR322 DNA and heated 10 min at 65°C. After ligation (as described in the previous section) and transformation into strain M94, trimethoprim-resistant clones were selected. Plasmid pCV27 was isolated from one such clone, CV628, that was resistant to ampicillin and sensitive to tetracycline.

Gel Electrophoresis

Agarose gels for analytical (14) and preparative (6) purposes have been

described. Phage λ DNA digested with EcoRI or HindIII was included in analytical gels to provide markers for molecular weight determinations.

Polyacrylamide gels (20 x 40 x 0.15 cm) for separation of double-stranded DNA fragments were run as described (15) using 90 mM Tris-borate, 2.5 mM EDTA, pH 8.3 (90 mM TBE buffer). Gels capable of resolving separated strands were run as described (16) except that 25 mM methyl mercuric hydroxide was used to denature the DNA (in 20 µl of E-buffer) (17) and electrophoresis was performed at 500V for 4-8 h without pre-electrophoresis. ^{32}P-labeled fragments were detected by autoradiography with Kodak XR-5 film and a Dupont Cronex lightning-plus intensifying screen. Unlabeled fragments were visualized by staining with 0.5 µg/ml ethidium bromide and illumination with UV light.

Gels for sequencing studies were 0.4 mm thick and 40 cm wide by 80 cm long. The notched plate was treated with siliconizing agent and the other plate cleaned with chromic acid. The plates were rinsed with distilled water and air dried before use. The gel mixture was as described (16) except that 60 or 90 mM TBE buffer was used (Table I). Acrylamide was purified from Eastman technical grade material using Bio-Rad AG 501-X8 mixed bed ion exchange resin. The conditions for electrophoresis are summarized in Table I. It was not necessary to pre-electrophorese the gels. After electrophoresis, the siliconized plate was removed and the gel (still stuck to the other plate) was covered with plastic film and autoradiographed as described above for 1-14 days at -20°C.

Isolation of DNA Fragments

DNA fragments were isolated from agarose gels by the method of Zain and Roberts (18) and from polyacrylamide as follows. A gel slice (20 x 5 x 1.5 mm) was macerated with a teflon-coated rod in a silicon-treated 12 x 75 mm glass tube. After addition of 1 ml of a solution containing 20 mM Tris-Cl, pH 8, and 2.5 mM EDTA, and incubation for 30 min at 45°C, or overnight at 4°C, the tube was centrifuged 10 min at 5,000 rpm. The extraction was repeated and the combined supernatants were applied to a 25 µl DE52 (Whatman) column (in a 1 ml disposable pipette tip plugged with silicon-treated glass wool) equilibrated with the same buffer. The last 0.5 ml was forced through the column under light air pressure. Double-stranded DNA was eluted with 0.4 ml of 2 M NaCl and precipitated with 1 ml of 95% ethanol in a 1.5 ml silicon-treated conical plastic centrifuge tube. Single-stranded DNA was eluted with 0.4 ml of 2 M NaCl containing 0.1 M NaOH and precipitated by the addition of 20 µl of 3 M sodium acetate, pH 5.5, followed by 40 µl of 1 N HCl and 1 ml of 95% ethanol. This procedure results in 80-95% recovery of the DNA. Fragments

Table I. Electrophoresis conditions for sequencing gels.

Number of nucleotides	Acrylamide[a] concentration	Number of loadings	Electrophoresis[b] conditions	Dye[c] migration
1-180	12.5%	1	3500V/ 4h	BP-2/3
10-250	6%	2	2200V/12h 2800V/ 4h	XC-3/4 BP-3/4
10-400	6%	3	2200V/13h 2800V/ 8h 2800V/ 4h	XC-7/8 XC-2/3 BP-3/4
20-400	4%	2	2200V/12h 2800V/ 4h	XC-4/5 BP-3/4
75-650	5%	3	2400V/14h 2800V/11h 3000V/ 6 h	XC-off gel XC-7/8 XC-2/3

[a]Gels containing 12.5% or 6% acrylamide contained 90 mM TBE. In gels with 4-5% acrylamide, 60 mM TBE was employed.

[b]The power did not exceed 60W. Distortion of the band pattern occurred at wattages higher than this.

[c]XC = xylene cyanol; BP = bromophenol blue. Fractions indicate the migration distance relative to the length of the gel.

for sequencing were re-precipitated from 50 µl of TE buffer with 1 ml of 95% ethanol and dissolved in 30 µl of TE buffer.

Labeling Procedures

Nick translation of DNA was carried out as described by Rigby et al. (19) except that the DNA [1-2 µg in 20 µl polymerase buffer (50 mM Tris-Cl, pH 8, 50 mM KCl, 5 mM $MgCl_2$, 5 mM dithiothreitol)] was incubated for 10 min at 68°C with 1 µl of a 5 x 10^{-6} dilution of a 1 mg/ml stock solution of DNaseI (Sigma, electrophoretically purified, stored in 0.15 M NaCl at -70°C in small aliquots). Following this, deoxynucleotide triphosphates [one or more α-^{32}P-labeled (Amersham, 400 Ci/mmol)] were added to 10-15 µM followed by 3 units of DNA polymerase I (New England BioLabs). The reaction was incubated at 37°C for 0.5-2 h. Unincorporated nucleotides were removed by ethanol precipitation. If the DNA was partially nicked to start with, the DNase step was omitted.

DNA fragments for sequencing were either labeled at the 5' end of strands by polynucleotide kinase (16) or more commonly at the 3' end of strands by treating with AMV reverse transcriptase (a gift from J. Beard). The following procedure for labeling 3' ends in molecules having 5' protruding ends was used. Ten pmols of plasmid DNA was cut with a restriction enzyme and, following ethanol precipitation, the DNA was dissolved in 20 µl of polymerase buffer. One hundred pmols of α-^{32}P-labeled deoxynucleoside triphosphate, [Amersham, 400 Ci/mmol; dATP for EcoRI and HinfI; dCTP for MspI and SalI (with unlabeled dTTP)] were dried under vacuum and dissolved in the reaction mixture. Twenty units (2 µl) of AMV reverse transcriptase were added and the mixture was incubated for 30 min at 37°C.

A new procedure was developed for labeling blunt-ended DNA fragments at their 3' ends. A DNA fragment created by endonuclease HaeIII was isolated from 15 pmols of plasmid DNA. It was treated with exonuclease III (New England BioLabs) (130 units in 200 µl of buffer containing 90 mM NaCl, 60 mM Tris-Cl, pH 8, 4 mM $MgCl_2$, 4 mM DTT) for 10 min on ice to remove about 5 nucleotides from the 3' ends (20). The volume was adjusted to 400 µl with TE buffer and 400 µl of TE-saturated phenol were added to stop the reaction. The aqueous phase was extracted 2 times with chloroform/isoamyl alcohol (24:1) and the DNA precipitated with 40 µl of 3 M sodium acetate, pH 5.5, and 1 ml of 95% ethanol. The precipitate was dissolved in polymerase buffer and labeled by treatment with AMV reverse transcriptase, α-^{32}P-dGTP (5 µM) and unlabeled dATP, dTTP, dCTP (each at 15 µM). In cases where more than 60% of the label was incorporated after 10 min, incubation was continued for an additional 20 min with 15 µM unlabeled dGTP to ensure that all labeled molecules were filled out to their ends. This labeling procedure has also been applied to blunt-ended fragments created by HincII. It should be applicable to any blunt-ended fragment.

DNA Sequencing Reactions

The chemical method of Maxam and Gilbert (16) was used with the following modifications. To the G reaction, only 0.5 µl of dimethylsulfate was added. Three µl of pyridinium formate were added to the A+G reaction. For fragments having 100-300 bases, the reaction times were 5 min at 20°C for the G, C and C+T reactions, and 25 min at 37°C for the A+G reaction. For fragments containing greater than 300 nucleotides, the reaction times were 10 min on ice for the G, C and C+T reactions and 10 min at 37°C for the A+G reaction. For long fragments, it was important to place the tubes on dry ice immediately after adding the stop solution and ethanol, and to process the samples through

the ethanol precipitation steps quickly. The piperidine cleavage was done in the original 1.5 ml plastic tubes with 50 µl of piperidine. Sealing the tubes with conformable tape was found to be unnecessary. After the reaction, the samples were transferred to another tube (otherwise it was impossible to completely remove the piperidine), 50 µl of 0.3 M sodium acetate and 0.4 ml of 95% ethanol were added, and the tubes centrifuged after incubation for 10 min on dry ice. The precipitates were rinsed with 1 ml of ethanol and dried for 15 min under vacuum before dissolving in a solution containing 98% formamide, 0.1% bromophenol blue and 0.1% xylene cyanol. Oligonucleotides as small as a trinucleotide are retained by this procedure. Ethanol precipitation rather than lyophilization was employed because it was faster.

RESULTS

Plasmids Conferring Resistance to Trimethoprim

Clones 04-46 and 14-37 of the Clarke-Carbon *E. coli* colony bank (11) are resistant to trimethoprim because they have multicopy plasmids (pLC446 and pLC1437) that contain the *E. coli* structural gene for DFRase (6). Plasmids pLC1437 and pLC1437a were recovered from different single colonies from isolate 14-37 of the colony bank. The relationship between plasmids pLC446, pLC1437 and pLC1437a was established by comparing sites at which restriction endonucleases cleave these DNAs (Fig. 1). The maps in Fig. 1 were constructed from data of the sort illustrated in Fig. 2. These results establish that a large fraction of the bacterial DNA insert (about 8.5 kb) is identical in the three plasmids. The sizes of these plasmids (determined by adding up the sizes of restriction fragments) were 19.5, 15.0 and 24.3 kb for pLC1437, pLC1437a and pLC446, respectively.

The Clarke-Carbon bank contains DNA from *E. coli* strain CS520 attached to plasmid ColE1. For our experiments, it was important to have cloned DNA from other strains. Accordingly, DNA from *E. coli* strain RSO was treated with endonuclease BamHI and ligated to BamHI-treated plasmid pBR322. Following transformation into strain M94, selection was made for resistance to trimethoprim. A restriction map of the plasmid isolated in this selection, pCV27, is also shown in Fig. 1. Treatment of plasmid pCV27 with endonuclease BamHI yields an 8.3 kb fragment of bacterial DNA that is identical or very similar to a fragment cut from plasmid pLC446 by the same enzyme. The specific activity of DFRase from strain CV628 (M94/pCV27), determined as described by Sheldon and Brenner (7), was 32 times that of the parent strain lacking the plasmid.

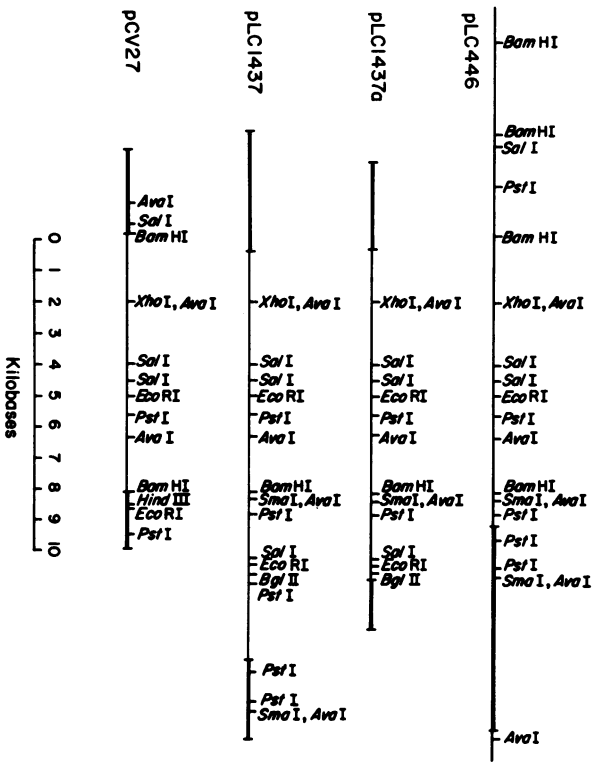


Figure 1. Restriction endonuclease cleavage site maps of plasmids pLC446, pLC1437a, pLC1437, and pCV27. The heavy portion corresponds to DNA of the cloning vector (pBR322 for pCV27; ColE1 for the others). Cleavage sites for restriction endonuclease BglII were not determined on plasmid pLC446.

Location of fol on Plasmid pLC1437a

fol was located relative to the restriction map of plasmid pLC1437a by determining the trimethoprim resistance of clones containing different portions of this plasmid. Four separate cloning experiments were carried out in which a fragment of plasmid pLC1437a was subcloned into plasmid pBR322 and the remaining portion of plasmid pLC1437a was resealed and cloned. Results with BglII/BamHI-generated clones and with PstI-generated clones (Table II, lines 1, 2, 6 and 7) indicate that fol is to the left of both PstI sites. On the other hand, the experiments with the two BamHI/SalI fragments (Table II, lines 4 and 5) demonstrate that fol is to the right of the pair of adjacent SalI sites. Therefore, fol lies within a 1.1 kb stretch between the right-most SalI site and the leftmost PstI site as indicated in Table II.

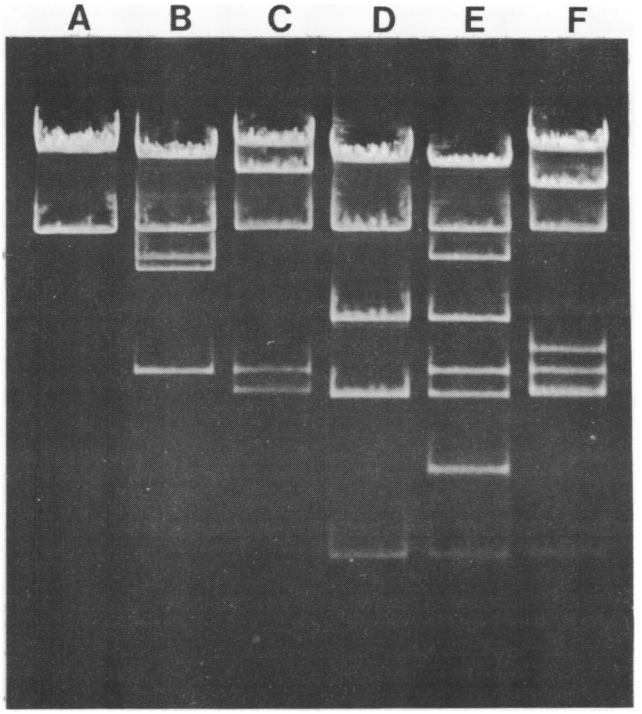


Figure 2. Agarose gel electrophoresis of plasmid DNAs pLC1437a (lanes A and D), pLC1437 (lanes B and E), and pLC446 (lanes C and F) digested with restriction enzymes PstI (lanes A, B and C) or PstI and SalI (lanes D, E and F). Electrophoresis was done for 2.5 h at 200V on a 1% agarose gel containing 90 mM TBE buffer.

Consistent with this interpretation is the fact that neither of the two hybrid plasmids derived by EcoRI cleavage (Table II, lines 3 and 8) conferred resistance to trimethoprim.

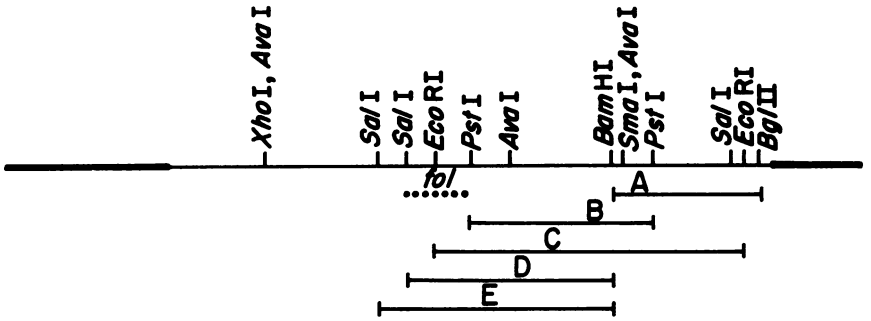
Nucleotide Sequence of Regions Flanking and Within fol

A detailed endonuclease cleavage site map of the DNA in the region containing fol was constructed (Fig. 3). In determining the sites at which various endonucleases act, it was convenient to use DNA labeled by nick translation and to detect fragments on gels by autoradiography. With such a procedure, only small amounts of DNA were required and small fragments were readily detected. The results of an experiment of this type are illustrated in Fig. 4A. In the nick-translation procedure described here, DNase I was inac-

Table II. Location of fol on plasmid pLC1437a.

Fragment cut from pLC1437a with enzyme(s)[a]	Size of fragment added or removed[b]	Cloning vehicle[c]	New plasmid designation	Sensitivity of transformants to trimethoprim[d]
Bgl/Bam (A)	+2.6	pBR322	pCV42	Trm^s
Pst (B)	+3.2	pBR322	pCV43	Trm^s
Eco (C)	+5.5	pBR322	pCV40	Trm^s
Bam/Sal (D)	+3.6	pBR322	pCV36	Trm^r
Bam/Sal (E)	+4.1	pBR322	pCV37	Trm^r
Bgl/Bam (A)	-2.6	pLC1437a	pCV39	Trm^r
Pst (B)	-3.2	pLC1437a	pCV38	Trm^r
Eco (C)	-5.5	pLC1437a	pCV41	Trm^s

[a] A,B,C,D, and E refer to fragments shown below.



[b] A plus sign indicates that a fragment cut out from pLC1437a was cloned on pBR322. A minus sign indicates that the plasmid remaining after removal of the indicated fragment from pLC1437a was cloned. Sizes are in kilobases.

[c] Strain JA200 or M94 served as recipient in transformations. Selection was made for ampicillin or tetracycline-resistance in experiments involving pBR322 and for colicin resistance when pLC1437a was the cloning vehicle.

[d] Ability (Trm^r) or inability (Trm^s) to grow in media containing 2 μg/μl trimethoprim.

tivated after only a small number of nicks were introduced into the DNA. The labeled DNA obtained in this way was usually sufficiently intact to allow strand separation of fragments in excess of 1 kb in length (Fig. 4B). It was

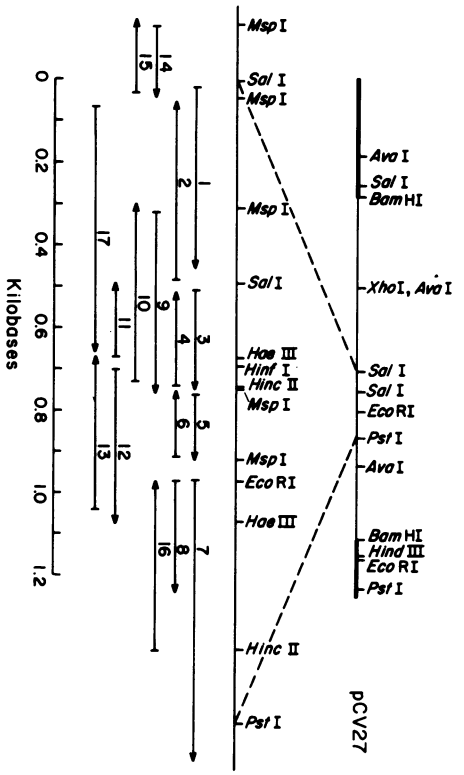


Figure 3. Sequencing strategy for the fol region. Fragments used in sequencing were derived from plasmid pCV27. The extent of sequencing in each case is indicated by the length of the arrow. Arrows point away from the end that was labeled. Sequences 1 and 2 were determined after 3' end labeling and strand separation of a SalI 500 base pair fragment. Sequences 3, 4, 5 and 6 were determined after 3' end labeling and strand separation of fragments derived from MspI cleavage of an AvaI/SalI 1.8 kb fragment. Sequences 7 and 8 were determined after 3' end labeling (sequence 7) or 5' end labeling (sequence 8) of EcoRI-cleaved pCV27 DNA followed by cleavage with AvaI. Sequences 9 and 10 were determined after 3' end labeling and strand separation of MspI fragments derived from an AvaI 4.3 kb fragment. Sequence 11 was determined after 3' end labeling of HinfI fragments derived from an AvaI 4.3 kb fragment followed by cleavage with SalI. Sequences 12 and 13 were determined after 3' end labeling and strand separation of a HaeIII 409 bp fragment. Sequences 14 and 15 were determined after 3' end labeling and strand separation of a 165 bp MspI fragment. Sequence 16 was determined after 3' end labeling of a 635 bp HincII fragment followed by cleavage with EcoRI. Sequence 17 was determined after 3' end labeling of a 1.6 kb SalI/PstI fragment (derived from a partial digestion). For the sake of clarity, only the restriction endonuclease cleavage sites relevant to this analysis have been included on this map.

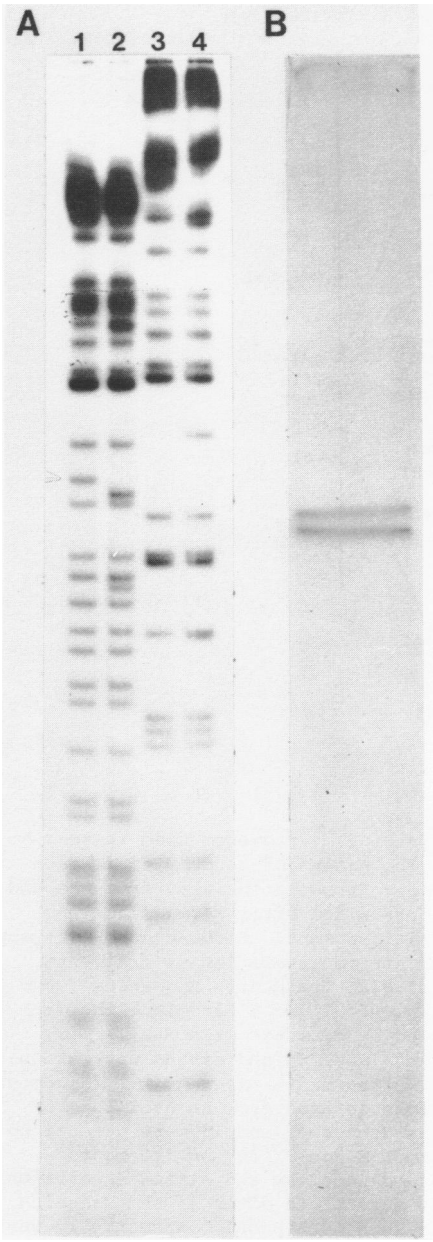


Figure 4. Acrylamide gel electrophoresis of nick-translated DNA fragments. (A) A 8.3 kb DNA fragment derived from BamHI cleavage of plasmid pCV27 was labeled and then digested with restriction endonucleases. Lane 1, MspI; lane 2, MspI/SalI; lane 3, HaeIII; lane 4, HaeIII/SalI. DNA samples contained about 5 x 10^5 cpm. Electrophoresis was performed at 400V for 5 h on a 20 x 40 x 0.15 cm polyacrylamide gel containing 6% acrylamide, 0.3% bis-acrylamide, and 90 mM TBE buffer. The bands were visualized by autoradiography for 2 h at room temperature. (B) Strand separation of a nick-translated DNA fragment. A 1.8 kb AvaI/SalI fragment was nick-translated and denatured in 25 mM methyl mercuric hydroxide. Five x 10^5 cpm was applied to a 0.15 x 30 x 40 cm polyacrylamide gel containing 4.5% acrylamide and 0.6% bis-acrylamide and 90 mM TBE buffer. Electrophoresis was done for 30 h at 300V and the gel was autoradiographed for 1 h.

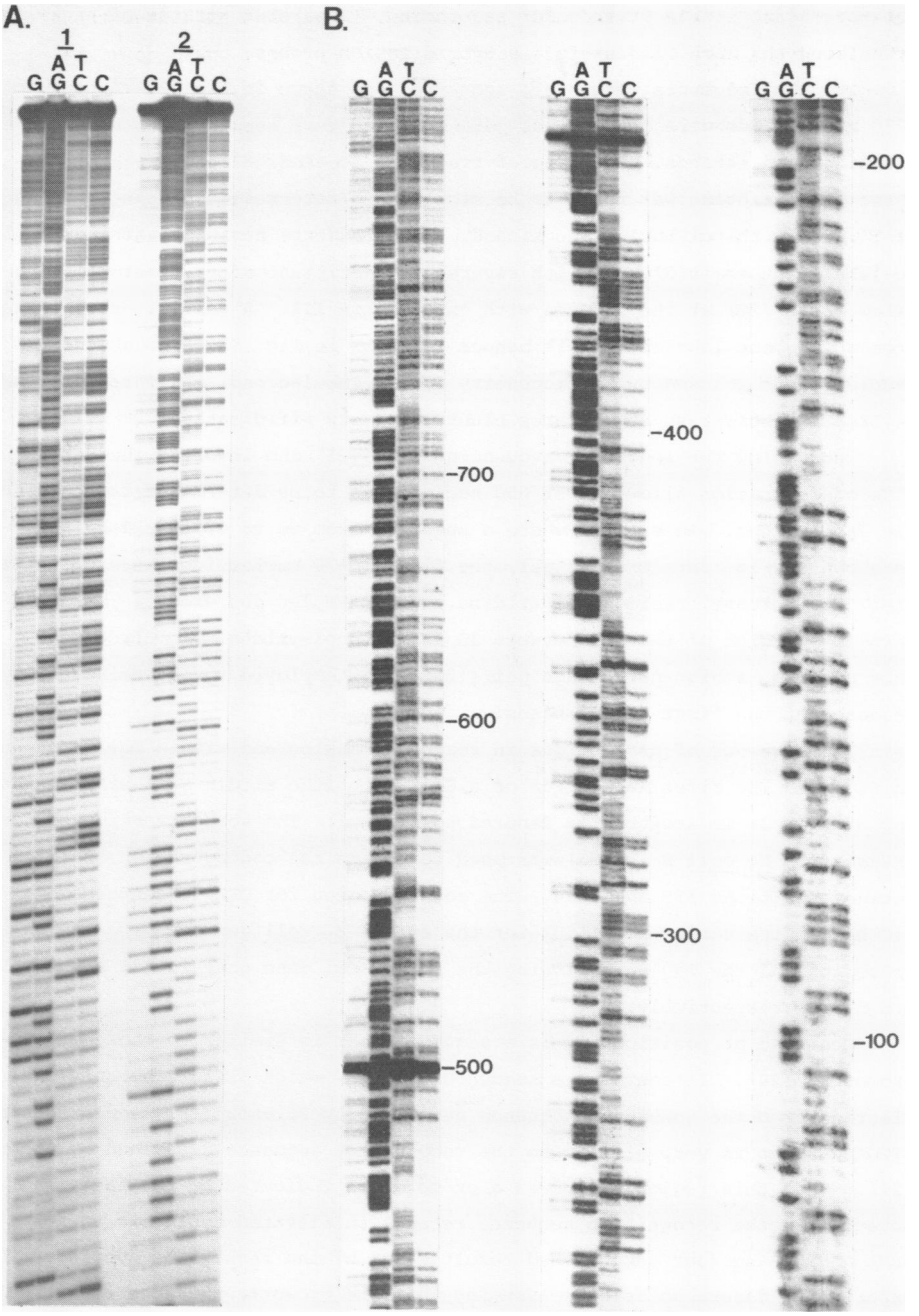
convenient to use nick-translated DNA to determine which fragments could be separated into single strands for sequencing. Separated strands of nick-translated DNA were also useful as hybridization probes.

The DNA fragments which were sequenced are shown in Fig. 3. A total of 1750 nucleotides were determined, 1450 of which have been confirmed by sequencing both strands. Labeling of fragments at their 3' ends with AMV reverse transcriptase was found to be simpler and more reliable than labeling at 5' ends with polynucleotide kinase. DNA fragments having blunt ends could be labeled at their 3' ends with reverse transcriptase after first removing a few bases (20) at the 3' ends with exonuclease III. A gel pattern obtained from a fragment labeled in this manner is shown in Fig. 5A. The absence of doublet bands (caused by heterogeneity at the labeled end) indicates that AMV reverse transcriptase regenerates blunt ends very efficiently.

Increasing the length of sequencing gels (21) and lowering the acrylamide concentration allowed over 650 nucleotides to be determined from a single labeled end. An example where a sequence from 80 to 680 could be read is shown in Fig. 5B. On a 4% gel, the first 10-20 nucleotides were not visible. Furthermore, residual piperidine in the samples obscured the band pattern in a region of the gel between 30 and 50 nucleotides. To circumvent this problem, a high percentage gel (12.5%) was employed to determine the sequence of the first 30-100 bases.

The sequence of nucleotides in regions flanking and within *fol* is shown in Fig. 6. The first nucleotide of a *Sal*I site (the middle *Sal*I site at the top of Fig. 3) is arbitrarily denoted position 1. The sequence reported for DFRase from *E. coli* B (22,23) was used to locate the coding region of *fol*, between positions 558 and 1034. The coding region for DFRase contains 53% G:C base pairs versus about 50% for the entire *E. coli* genome. The 200 bp preceding and the 150 bp following the structural gene contain 42% and 50% G:C pairs, respectively.

Centered at position 514 is a sequence that is similar to those of known promoters (24). It contains a sequence TATAGTG, which differs by only one nucleotide from the consensus sequence described by Pribnow (25), and a sequence GTCGACGA that is very similar to the recognition sequence described by others (24). That this region is indeed a promoter is indicated by the fact that a mutation in the recognition sequence results in elevated synthesis of *fol* mRNA and DFRase (our unpublished results) and by the results of hybridization experiments described below. Assuming that the promoter sequence is correctly identified, mRNA synthesis is expected to start about 10 nucleotides down-


A.
1
2
G A T
G C C
B.
-700
-600
-500
-400
-300
-200
-100

stream from the center of the Pribnow sequence, at or near position 535. Between the beginning of the mRNA and the AUG codon that initiates translation (about 23 nucleotides) is a sequence TCGGGAAAT that could serve as a ribosome binding site (26).

Upstream from fol is another potential promoter located between positions 395 and 425. The alignment between this region and consensus sequences is reasonably good (6/7 in the Pribnow region, 7/12 in the recognition sequence). A mRNA initiated at this site would contain a potential ribosome binding site located at position 440 and a nearby GUG codon at which translation might initiate (26). The GUG codon is in the same reading frame as fol. Translation initiated at this codon would result in a 196 amino acid polypeptide containing the DFRase sequence at its C-terminal end (DFRase contains 159 amino acids). Results presented below suggest that if transcription is initiated at this putative promoter, it is at a relatively low frequency.

Within the 1200 bp sequence shown, one other region seems capable of encoding a polypeptide (between positions 175 and 363). This region, including a potential promoter (between positions 91 and 121), two ribosome binding sites (positions 164 and 172), with in-phase AUG codons (positions 175 and 181) and tandem UAA codons (positions 364 and 367), can potentially encode two polypeptides having 61 or 63 amino acids.

RNA Transcripts Initiated at the fol Promoter

RNA from a strain harboring a multicopy fol-containing plasmid was fractionated on a methyl mercuric hydroxide-agarose gel and bound to DBM-paper. Hybridization of a labeled HincII/EcoRI 225 bp fragment (totally within the DFRase structural gene) followed by autoradiography revealed a major transcript of about 550 nucleotides (Fig. 7, track C). Another major transcript of about 590 nucleotides was sometimes also present. It is not apparent in the sample analyzed in Fig. 7, track C, but was present in other isolates of that RNA and was visible in RNA from strain RS35 (Fig. 7, track B). In addi-

Figure 5. Autoradiograms from which part of the sequence of fol was determined. (A) A 409 bp HaeIII fragment (Fig. 3) was labeled at the 3' end with AMV reverse transcriptase following treatment with exonuclease III and the strands were separated (lane 1, slow strand; lane 2, fast strand). A 12.5% gel was used (Table I). (B) A 1.6 kb SalI/PstI fragment (Fig. 3) was labeled at the SalI site with the aid of AMV reverse transcriptase. The spacing of bands up to nucleotide 680 was wide enough to allow an unambiguous sequence to be read. A 5% gel was used (Table I). The heavy band at position 500 is caused by partial cleavage by endonuclease SalI at the site located there.

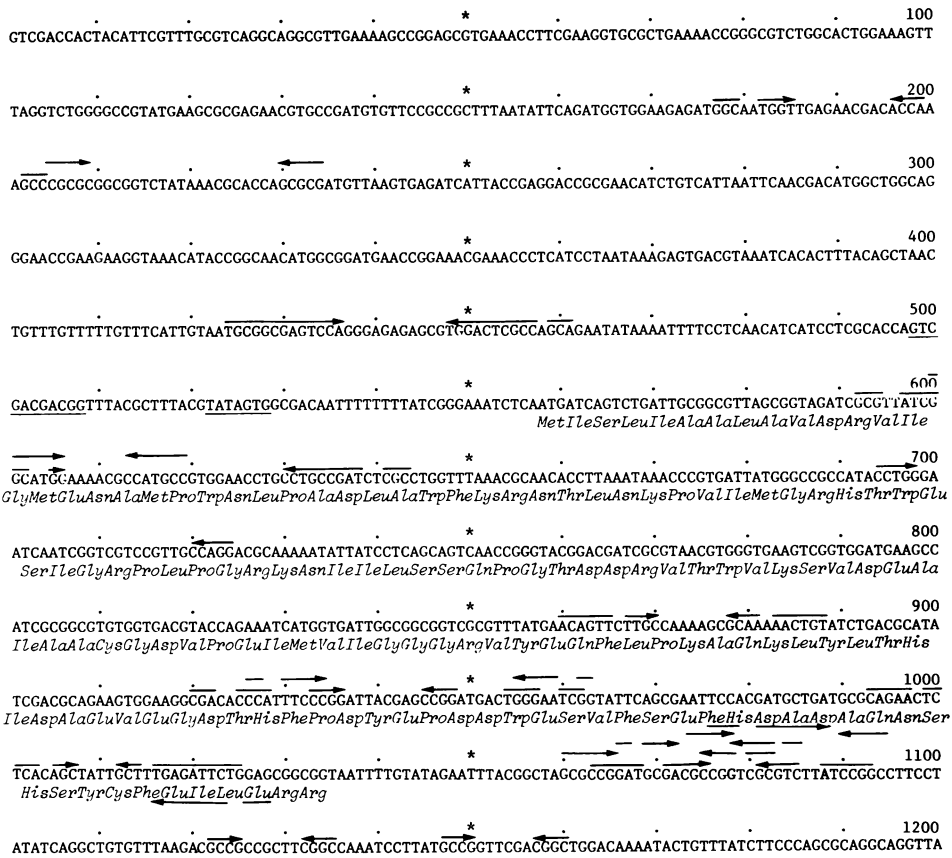


Figure 6. Nucleotide sequence of the E. coli K12 fol region from strain RSO. The sequence goes from top to bottom, left to right, in a 5'→3' direction, with only one DNA strand being shown. The amino acid sequence of DFRase predicted from the nucleotide sequence is shown. The proposed location of the fol promoter is underlined. Arrows indicate regions of dyad symmetry, identified using the computer program of Queen and Korn (27). Only those dyad symmetries are included which have a loop distance less than 30 nucleotides and a free energy of -8 kcal/mol or less, as calculated by the method of Borer et al. (28). Destabilization due to loops and mismatches (29) was included in the calculation and G:U pairs were assigned a value of zero kcal/mol.

tion to these, fainter bands corresponding to transcripts of about 720, 900, 1000, 1100, 1300, and 1400 nucleotides were observed. On the other hand, there was relatively little hybridization (about 15-fold less based on densitometry scans) when a 103 bp AluI/SalI fragment (located between positions

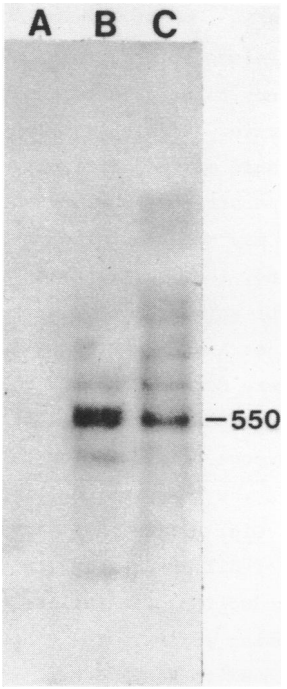


Figure 7. RNA transcripts from the fol region. RNA samples, prepared as described previously (6), were separated on a methyl mercuric hydroxide-agarose gel and transferred to DBM-paper (30). A 225 bp HincII/EcoRI fragment (located between positions 751 and 976, Fig. 6) was labeled by nick translation to about 10^7 cpm/µg. Hybridization was carried out with about 5 x 10^6 cpm of DNA as described (31) except that the temperature was 50°C. The filter was autoradiographed for two days with a Dupont Cronex lightning plus intensifying screen. A, RNA from strain RSO (40 µg); B, RNA from strain RS35 (fol-60) (40 µg); C, RNA from strain CV652 (BL322/pCV27) (20 µg). Approximate sizes of the RNA bands were determined from a standard curve, using E. coli ribosomal RNAs as standards. The two small RNA bands in track B were not seen in other experiments with the same RNA.

395 and 498 in Fig. 6) was used as a probe (data not shown). These results are consistent with the proposed location of the fol promoter and indicate that in this plasmid, the region to the left of fol is not transcribed to a great extent. They also show that the longer fol transcripts do not derive from the region to the left of the fol promoter, and must therefore be longer at the 3' end. Consistent with this, probes derived from the region beyond the 3' end of fol hybridize only to the larger transcripts (data not shown).

When this experiment was repeated with RNA from a wild-type strain, no significant hybridization was observed (Fig. 7, track A). DFRase amounts to less than 0.01% of the total cell protein of E. coli and the procedures used were apparently not sufficiently sensitive to detect fol mRNA in this strain. On the other hand, fol RNA from strain RS35 (fol-60) was sufficiently abundant to be readily recognized in this experiment (Fig. 7, track B). Furthermore, the number and sizes of fol mRNAs from strain RS35 (fol-60) and BL322/pCV27 were very similar, suggesting that mutation fol-60 does not grossly affect the point of initiation or termination of fol mRNA and that transcription of plasmid genes does not affect transcription of fol carried on plasmid pCV27.

DISCUSSION

Since our main interest is in understanding mechanisms of drug resistance brought about by mutation, it is of paramount importance to be able to analyze sequences from both mutant and wild-type cells. A strategy involving polypeptide sequencing is not attractive in this case because of the difficulty in isolating a suitable amount of enzyme from wild-type strains, but more importantly because of the possibility that mutations may occur outside of the structural gene. DNA sequencing, on the other hand, is both fast and accurate and is equally applicable to parent and mutant strains.

The amino acid sequences of DFRases from E. coli strains RT500 and MB1428 have been determined (22,23). These strains were derived from E. coli B by selection with trimethoprim or methotrexate and produce high levels of DFRase. A comparison of these sequences with the sequence predicted for the E. coli K12 wild-type enzyme shows differences between the three at positions 37, 87, 118, 142 and 154: RT500, Asp, Asn, Gln, Asp, Glu; MB1428, Asp, Asn, Glu, Asn, Lys; K12, Asn, Asp, Glu, Asp, Glu. These few differences may be due to strain differences (K12 versus B), to the introduction of mutations during the selection of drug resistance, or to sequencing errors. The RT500 sequence has recently been revised at positions 37 and 87 to Asn and Asp, respectively (personal communication from David Stone).

The major RNA transcripts of the fol region are approximately 550 and 590 nucleotides long (Fig. 7). If these transcripts initiate at the fol promoter, both must have their 3' ends in a region of about 100 nucleotides following the DFRase translation stop codon. The existence of longer transcripts containing up to 1400 nucleotides suggests either that RNA polymerase sometimes reads through past one or more transcription terminators or that 1400 nucleotides are transcribed and then processed, yielding mainly fragments 550 and 590 nucleotides long. The region to the right of the TAA that signals termination of fol translation (position 1050 to 1160) contains many G:C rich palindromes (Fig. 6). None of these is followed by four or more thymine residues as expected for a strong rho-independent transcription stop signal (24). However, several of the palindromes, including those centered at positions 1070 and 1128, resemble rho-dependent terminators (24). Finally, some of these palindromic sequences may pair in a manner reminiscent of known RNA processing sites (32-34).

ACKNOWLEDGEMENTS

This work was supported in part by funds from the Cornell University Agricultural Experiment Station. We thank Robert Sheldon and William Studier for strains, Joseph Beard for AMV reverse transcriptase, and Elizabeth Keller and Ray Wu for helpful discussions and reviews of the manuscript.

REFERENCES

1. Bertino, J., Johns, D. (1972) in Cancer Chemotherapy (I. Bradsky, ed.) Vol. 2, pp. 9-22, Grane and Stratton, NY.
2. Sirotnak, F.M. (1970) Genetics 65, 391-406.
3. Kellems, R.E., Alt, F.W., Schimke, R.T. (1976) J. Biol. Chem. 251, 6987-6993.
4. Chang, S.E., Littlefield, J.W. (1976) Cell 7, 391-396.
5. Alt, F.W., Kellems, R.E., Bertino, J.R., Schimke, R.T. (1978) J. Biol. Chem. 253, 1357-1370.
6. Smith, D.R., Calvo, J.M. (1979) Mol. Gen. Genet. 175, 31-38.
7. Sheldon, R., Brenner, S. (1976) Mol. Gen. Genet. 147, 91-97.
8. Breeze, A.S., Sims, P., Stacey, K.A. (1975) Genet. Res. (Camb.) 25, 207-214.
9. Sheldon, R. (1977) Mol. Gen. Genet. 151, 215-219.
10. Tritz, G.J., Matney, T.S., Gholson, R.K. (1970) J. Bacteriol. 102, 377-381.
11. Clarke, L., Carbon, J. (1976) Cell 9, 91-99.
12. Miller, J.H. (1972) in Experiments in Molecular Genetics. Cold Spring Harbor Press, NY.
13. Barnes, W. (1977) Science 195, 393-394.
14. Wu, R., Jay, E., Roychoudhury, R. (1976) in Methods in Cancer Research (H. Busch, ed.) Vol. 12, pp. 87-176, Academic Press, NY.
15. Maniatis, T., Jeffrey, A., van de Sande, H. (1975) Biochemistry 14, 3787-3794.
16. Maxam, A.M., Gilbert, W. (1979) in Methods in Enzymology (in press).
17. Bailey, J.M., Davidson, N. (1976) Anal. Biochem. 70, 75-85.
18. Zain, B.S., Roberts, R.J. (1978) J. Mol. Biol. 120, 13-31.
19. Rigby, P.W.J., Dieckmann, M., Rhodes, C., Berg, P. (1977) J. Mol. Biol. 113, 237-251.
20. Donelson, J.E., Wu, R. (1972) J. Biol. Chem. 247, 4661-4668.
21. Yang, R.C.A., Wu, R. (1979) Virology 92, 340-352.
22. Stone, D., Phillips, A.W., Burchall, J.J. (1977) Eur. J. Biochem. 72, 613-624.
23. Bennett, C.D., Rodkey, J.A., Sondey, J.M., Hirschman, R. (1978) Biochemistry 17, 1328-1337.
24. Rosenberg, M., Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
25. Pribnow, D. (1975) Proc. Natl. Acad. Sci. 72, 784-788.
26. Steitz, J.A. (1979) in Biological Regulation and Development (R.F. Goldberger, ed.) Vol. 1, pp. 349-399, Plenum Press, NY.
27. Queen, C., Korn, L.J. (1979) in Methods in Enzymology, vol. 65 (in press).
28. Borer, P.N., Dengler, B., Tinoco, I., Jr., Uhlenbeck, O.C. (1974) J. Mol. Biol. 86, 843-853.
29. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., Gralla, J. (1973) Nature New Biol. 246, 40-41.
30. Alwine, J.C., Kemp, D.S., Parker, B.A., Reisev, J., Renart, J., Stark, G.R., Wahl, G.M. (1979) in Methods in Enzymology (in press).

31. Wahl, G.M., Stern, M., Stark, G.R. (1979) Proc. Natl. Acad. Sci. 76, 3683-3687.
32. Rosenberg, M., Kramer, R.A. (1977) Proc. Natl. Acad. Sci. 74, 984-988.
33. Oakley, J.L., Coleman, J.E. (1977) Proc. Natl. Acad. Sci. 74, 4266-4270.
34. McConnell, D.J. (1979) Nucleic Acids Res. 6, 3491-3503.