Volume 15 Number 20 1987 Nucleic Acids Research

Nucleotide sequence of the *Dde*I restriction-modification system and characterization of the methylase protein

Laura A.Sznyter, Barton Slatko, Laurie Moran, Kimberly H.O'Donnell* and Joan E.Brooks

New England Biolabs, Beverly, MA 01915, USA

Received July 20, 1987; Revised and Accepted September 21, 1987 Accession no. Y00449

ABSTRACT

The *Dde*I restriction-modification system was previously cloned and has been maintained in *E. coli* on two separate and compatible plasmids (1). The nucleotide sequence of the endonuclease and methylase genes has now been determined; it predicts proteins of 240 amino acids, M_r=27,808, and 415 amino acids, M_r=47,081, respectively. Inspection of the DNA sequence shows that the 3' end of the methylase gene had been deleted during cloning. The clone containing the complete methylase gene was made and compared to that containing the truncated gene; only clones containing the truncated form support the endonuclease gene in *E. coli*. *Bal*-31 deletion studies show that methylase expression in the *Dde* clones is also dependent upon orientation of the gene with respect to pBR322. The truncated and complete forms of the methylase protein were purified and compared; the truncated form appears to be more stable and active *in vitro*. Finally, comparison of the deduced amino acid sequence of M. *Dde*I with that of other known cytosine methylases shows significant regions of homology.

INTRODUCTION

Restriction-modification systems are found in a wide variety of prokaryotic organisms where they protect the host from infection by foreign DNA (2, 3, 4). The process of cloning these systems has benefitted from recent advances. First, the widespread use of methylase selection, in which digestion of random clones with a restriction endonuclease directly selects those that express the sequence-specific protective methylase, has resulted in a large number of methylase clones (M.*Hha*I, 5; M.*Msp*I, 6; M.*Bsp*RI, 7; M.*Dpn*II, 8). Since the methylase(*m*) and endonuclease(*r*) genes are often closely linked, they are frequently cloned together (*Bsu*RI, 9; *Taq*I, 10; *Dpn*II, 11).

A second advance has been the awareness of a series of recently-identified restriction systems in *E. coli*, specific for modified DNA. It has been found that DNA introduced into *E. coli* K12 from a foreign host, carrying a methylation pattern different from that of *E. coli*, often transforms or infects with low efficiency (12, 13). At present two restriction systems, McrA and McrB, specific for DNA containing methylcytosine, are the best characterized (13, 14, 15); there is evidence that at least two more methylcytosine-

targeted systems exist in *E. coli* K12 strains (E. Raleigh, pers. comm.). In addition, one system, Mrr, specific for DNA containing methyladenine, has recently been identified (16). These systems all have some sequence specificity; none of them restrict DNA modified by methylases normally found in *E. coli* K12. The degree of restriction appears to be proportional to the degree of methylation; therefore, the use of *E. coli* strains deficient in one or more of these functions has been essential for the successful cloning of many of the methylase genes.

The *Dde*I restriction-modification system was cloned into *E. coli* in two steps (1). First, methylase selection was used to obtain a 3.0kb *Hind*III fragment cloned into pBR322 (pDdeM3.0a), which contains the *Dde*I methylase gene. A subclone pDdeM1.6, which has a higher level of methylase expression, was constructed. However, it was not possible to clone the complete system into *E. coli* in one step. A second step was needed to clone the endonuclease gene. The *r* gene was located on a *Bam*HI digest of the *Dde*I genomic DNA by Southern hybridization with an oligonucleotide probe specific for the 5'-terminus of the *r* gene. A 2.3kb fragment large enough to contain the *r* gene hybridized to the probe and was cloned into pACYC184. In the presence of the "hyperactive" methylase subclone pDdeM1.6 , the *r* gene was introduced (Figure 1). In this way the *Dde*I system is maintained in *E. coli* on two separate and compatible plasmids.

Surprisingly, the pDdeR2.3 plasmid can be transformed into cells containing pDdeM3.0a or b at the same efficiency as those containing pDdeM1.6. However, upon subsequent growth these transformants prove inviable; therefore, it became clear that the high level of methylase expression is necessary for successful *r* gene maintenance. However, it remained unclear why our particular subclone should express more methylase than clones containing the progenitor plasmid. It was hoped that determining the sequence of the *Dde*I system would explain this phenomenon, as well as improve our understanding of how the *Dde*I system is arranged and regulated.

MATERIALS AND METHODS

Bacterial Strain, Plasmids, and Phage

E. coli strain ER1467(*mcr*B-) has been described previously (1, 13), and was grown in LB Medium (see ref. 17).

The plasmids pUC18 and pUC19 (18) were used in several subcloning experiments. Plasmids pDdeM1.6, pDdeR2.3 and pDdeM3.0a and their construction have been previously described (1). Plasmid DNA was isolated by the cleared lysate method (19), then subjected to CsCl/ethidium bromide ultracentrifugation (see ref. 17). Small-scale plasmid preparations were made using the alkaline lysis method (see ref. 17).

Phage stocks of λ_{vir} (New England Biolabs collection) were prepared according to Maniatis (17) and all phage dilutions were done in SM solution (17).

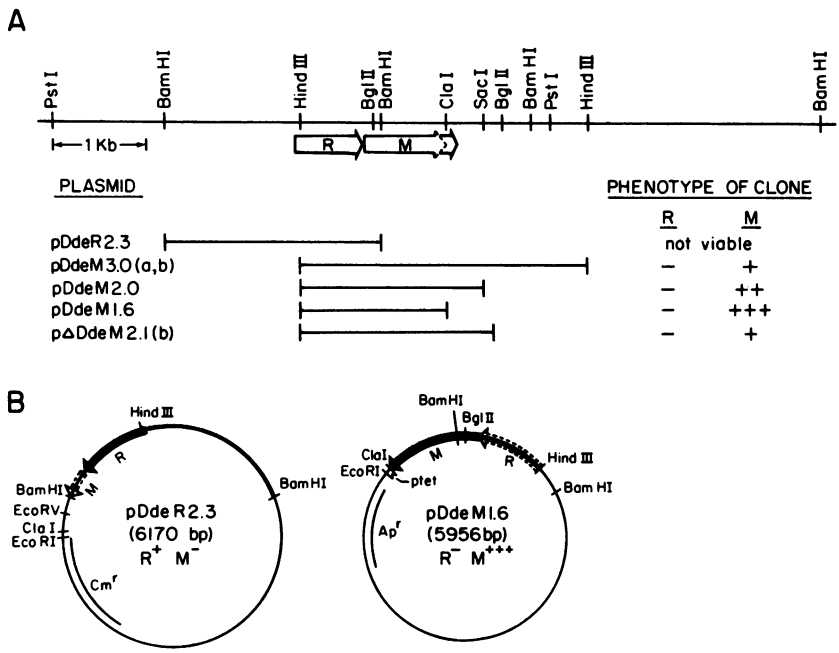


Figure 1. Restriction map of *Desulfovibrio desulfuricans* chromosome and *Dde*I clones. (A): Restriction map of *D. desulfuricans* genome in the region containing the *Dde*I system and derivative plasmids. *Arrows* indicate orientation of the genes on the chromosome. *Broken arrow head* indicates the carboxyl terminus of the methylase from pDdeM1.6. pDdeM3.0a, pDdeM3.0b, pDdeM2.0, pDdeM1.6, pΔDdeM2.1b are pBR322 derivatives carrying the *m* gene; pDdeM3.0a and pDdeM3.0b are identical except the insert is present in the opposite orientations relative to pBR322; pDdeR2.3 is a pACYC184 derivative carrying the *r* gene alone. +, ++, +++ refers to the relative amounts of methylase produced by each clone. (B): Plasmids containing the viable *Dde*I restriction-modification system in *E. coli*. *Heavy black lines* indicate cloned regions of the chromosome. *Arrows* indicate orientation of the genes; *dotted arrows* indicate incomplete and therefore inactive forms of the genes [*r*≡endonuclease; *m*≡methylase; p*tet*≡tetracycline promoter of pBR322].

Plasmids carrying *Dde*I genes were designated in accord with the previously described nomenclature (20).

Cloning Enzymes and Techniques

All restriction endonucleases and other cloning enzymes were purified at New England Biolabs and used according to prescribed conditions. M.*Alu*I, M.*Taq*I, M.*Pst*I and M.*Eco*RI were purified at New England Biolabs and used as recommended except [methyl-^{3}H] S-adenosyl-L-methionine [12.8 Ci/mmole, New England Nuclear] was used

as the methyl donor. A 5 to 20-fold excess of methylase was used per assay; reactions were incubated for 30 minutes at the recommended temperature.

The $CaCl_2$ heat shock method (21) was used for all transformations.

Cloning of the Complete Methylase Gene

pDdeM2.0, which contains the complete *Dde*I *m* gene, was constructed as shown in Figure 1. Five μg of pDdeM3.0a was simultaneously digested with 40 units each of *Eco*RI and *Sac*I for one hour under recommended conditions. Following phenol and chloroform extraction, ethanol precipitation, and resuspension in 10 μl dH_2O, the ends were filled in with 10 units of the Klenow fragment from DNA Polymerase I in a final volume of 25μl, incubating at 37°C for 30 minutes in the recommended buffer. The plasmid was again extracted, precipitated, and resuspended in 100μl of recommended ligase buffer and ligated at 16°C overnight with 1000 units T4 DNA Ligase. The ligation mix was chloroform extracted and used to transform ER1467.

DNA Sequencing

The dideoxynucleotide chain-termination procedure of Sanger et al.(22) was used for DNA sequencing experiments; [α-^{35}S] dATP or [α-^{32}P] dATP DNA sequencing reagents, primers and Klenow polymerase (New England Biolabs) were employed. Double-stranded templates were prepared as previously described (23, 24, 25). [α-^{32}P] dATP (3000 C_i/mmol) or [α-^{35}S] dATP (1200 C_i/mmol) (NEN) was used to label the DNA. Templates include pDdeM1.6, and pBR322 or pUC19 subclones carrying restriction fragments from pDdeM1.6, pDdeM2.0 or pDdeR2.3 (Figure 2). The appropriate pBR322 primers and pUC forward and reverse universal primers were utilized, and oligonucleotide primers complementary to regions within the 1.6 kb pDdeM1.6 insert were synthesized (New England Biolabs-Organic Synthesis Division) and used as internal primers. DNA sequences were resolved using both 8% standard and "wedge" polyacrylamide gels (26, 27) run in TBE buffer. Gels were fixed, dried, and processed as described in (28).

Computer Analysis of Nucleotide Sequence

A combination of programs from the University of Wisconsin (29), Cold Spring Harbor (30) and Align (31) were used for sequence analysis.

Nuclease *Bal*-31 Deletions

Deletions in plasmids pDde3.0a and pDde3.0b were constructed using nuclease *Bal*-31 as described below. Between each step, the DNA was extracted with phenol, then chloroform, and precipitated with two volumes of ethanol. Thirty-five μg of plasmid DNA was linearized with either a 10-fold excess of *Eco*RI or a 15-fold excess of *Eco*RV in a reaction volume of 700μl, under recommended conditions. After incubation for 2 hours the DNA was extracted, precipitated and resuspended in 35μl TE (10mM Tris, pH8.0; 1mM EDTA) and was then treated with 2.7 units of *Bal*-31 in a reaction volume

of 54μl under the specified conditions. Aliquots (8 μl) were removed at 1, 5, 10, 20, 30, 60, and 90 minutes, added to tubes containing 100μl H_2O and 100μl phenol, and vortexed to stop the reaction. After extraction and precipitation, the resuspended DNA was treated with 25 units of Klenow in a reaction containing 44mM each of dATP, dCTP, dGTP and dTTP (30 minutes, 37°C) in a final volume of 25μl. After extraction and precipitation, the plasmid was resuspended in 500μl ligase buffer and ligated overnight to a 50-fold molar excess of *Xho*I linker(pCCTCGAGG;NEB) with 10,000 units T4 DNA Ligase under recommended conditions. Following reextraction, precipitation and resuspension, the DNA was cleaved with 50 units of *Xho*I in a final volume of 100μl under prescribed conditions. After one hour the reaction was terminated by heating at 65°C for 10 minutes, the DNA was diluted to a final concentration of 10μg/ml in 500μl and religated with 10,000 units of T4 DNA Ligase under standard conditions.

Analysis of *Dde*I Endonuclease and Methylase Activity

In vitro endonuclease activity was determined as previously described (1). To determine *in vitro* methylase activity, ^{3}H-methyl incorporation was measured as previously described (1). Methylase units were quantified using the protection assay described previously (32) with the following modifications: 1 μg of pBR322 linearized with *Eco*RI was the methylation substrate; subsequent to methylation, the DNA was challenged with 30 units of *Dde*I endonuclease for 30 minutes under standard conditions. One unit of *Dde*I methylase is defined as the amount of enzyme which, in a 20 minute reaction at 37°C, is required to completely protect 1 μg of pBR322/*Eco*RI against *Dde*I endonucleolytic cleavage.

In vivo methylase activity was determined by passage of λ_{vir} phage through the test host as previously reported (1).

*Dde*I Methylase Purification

One liter of ER1467 cells containing either pDdeM1.6 or pDdeM2.0 was grown to a Klett of 90 (mid log phase) at 37°C. The cells were harvested by centrifugation and the pellets (1.5 grams and 1.4 grams respectively) were stored overnight at -70°C. The frozen pellets were thawed on ice and resuspended in 6 mls Sonication Buffer (SB; 10mM Tris pH 7.6, 10 mM β-mercaptoethanol, 0.1mM EDTA). All steps described below were done at 4°C. Lysozyme was added to a final concentration of 500 μg/ml, and then the cells were disrupted by sonication (two 10 second bursts). The cell lysate was brought to 0.05M NaCl and spun at 10,000 x g for 40 minutes. The supernatant was loaded onto a 1.5 x 3 cm Heparin-Sepharose (Pharmacia) column equilibrated and washed with SB containing 0.05M NaCl. The methylase was eluted with a 60 ml linear gradient from 0.05M to 1.0M NaCl. The methylase activity eluted at 0.6M NaCl. Active fractions (7.25 and 6.5 ml, respectively) were pooled and dialyzed against Buffer A (20mM Tris, pH 7.6; 10mM β-mercaptoethanol) containing 0.05M KCl.

TABLE 1. PURIFICATION OF THE TRUNCATED *Dde* I METHYLASE

Purification Step	Total Protein (mg)	Total Activity (Units)	Yield (%)	Specific Activity (U/mg)
Crude Extract	90	7500	100	83.3
Heparin Sepharose	4.35	1200	16	277.8
Mono Q	0.125	125	1.6	1000
polyCAT A	0.0125	100	1.3	8000

Each 8 ml methylase pool was adsorbed onto a Mono Q column (Pharmacia) equilibrated in Buffer A containing 0.05M KCl, and a 40 ml linear gradient 0.05M to 0.5M KCl was run. The methylase activity eluted at 0.13M KCl.

The active fractions (5 ml) from the Mono Q columns were dialyzed overnight against Buffer B (20mM KPO_4, pH 7.0; 10mM β-mercaptoethanol) containing 0.05M KCl, and then loaded onto a poly CAT A column (Pharmacia) equilibrated at 0.05M KCl in Buffer B and eluted with a 42 ml gradient from 0.05M to 0.5M KCl. The methylase activity eluted at 0.32M KCl as a single peak in each case.

The methylase activity of column fractions was determined via the ^{3}H-methyl incorporation assay; the yield and specific activity of each step was calculated using the methylase protection assay. Protein concentrations were determined via the dye binding assay of Bradford (33) standardized with bovine serum albumin. Table 1 shows purification and yield of *Dde*I methylase from pDdeM1.6.

RESULTS

DNA Sequence Analysis

The sequence of a 2256 base region including the *Dde*I restriction-modification system has been determined. The sequencing strategy is illustrated in Figure 2; the nucleotide sequence along with the predicted amino acid sequence of the methylase and endonuclease proteins is given in Figure 3.

In our previous work, the relative position and approximate gene boundaries of *r* and *m* were determined (Figure 1). The sequence shows that the *r* gene precedes the *m* gene with both being transcribed in the same direction. The arrangement of the *Dde*I system is therefore analogous to that of *Eco*RI (34,35) and *Bsu*RI (9).

The DNA sequence was examined for open reading frames. Only two are found that are large enough to encode the *r* and *m* genes. The first of these starts at an ATG at position 138 and terminates with a TAG at position 858. This is ascribed to

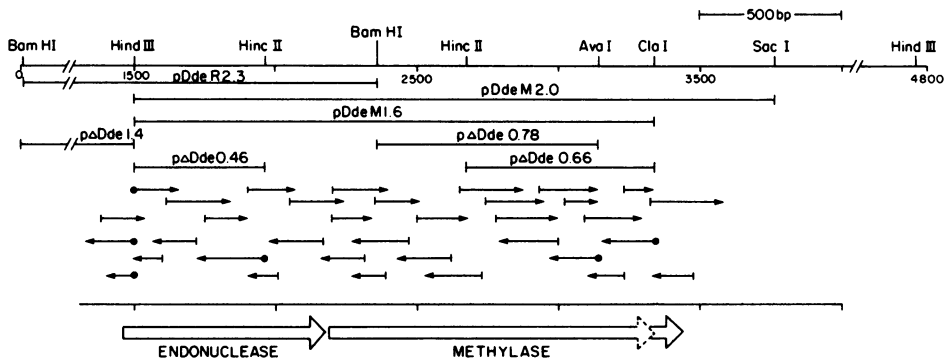


Figure 2. Fine structure restriction map and mode of sequencing the *Dde* I region. Restriction sites noted on the top line were used to generate subclones for sequencing. Fragments of the *Dde* I which were subcloned into pUC19 or pBR322 and used to generate sequence are shown in the 6 lines below the restriction map. The *small arrows* indicate the location, direction and length of sequence read in each set of reactions. Arrows beginning with circles (o→) correspond to sequencing reactions primed with universal primers; those with vertical lines (→) indicate those primed with unique primers. Each region was covered by sequence reactions in both directions. The amino terminal sequence of the *r* gene was confirmed by protein sequence. Location and orientation of the endonuclease and methylase genes are indicated by *large arrows*. The *broken arrow head* indicates the carboxyl terminus of M.*Dde*I'.

the *r* gene; its nucleotide sequence exactly predicts the first 19 amino acids found by sequencing the amino terminus of the endonuclease protein (1). The open reading frame would specify a 240 amino acid protein of 27,808 daltons. This molecular weight agrees well with the value obtained for the homogeneous endonuclease by SDS-PAGE(1).

Six bases downstream and in phase with the *r* gene is an ATG at position 864 which begins the second open reading frame, 1245 base pairs in length; this is attributed to the *m* gene. This open reading frame terminates with a TAA at position 2109. However, 54 bases downstream from the first ATG is a second ATG at position 918. From the first ATG a protein of 415 amino acids, molecular weight 47,081 would be made; the second ATG would give a 388 residue protein (M_r=43,736). Attempts to sequence the amino terminus of the purified methylase proved unsuccessful, so it is not possible at present to assign unambiguously the start of the *m* gene.

However, a series of *Bal*-31 deletion experiments suggest that the first ATG is actually the start of the *m* gene. Deletions terminating prior to position 864 result in clones that retain methylase activity, but a deletion terminating at position 909 (between the two potential starts) yields a complete methylase deficiency. This suggests that the

```
   1  CGAAAAGGAA AGGTTGATGA AGAACTGGTT AAATGTCGAA TATCAATGAA GGATTAGGAT      60

  61  TTATGTAGGG TTGACTCGTA GATACCTATA ATATAATGGA TCGGAATTTA TATAATATTG     120

                                M  K  A  A  T  D  Q  E  L  R  K  L
 121  TAGGAGTGGG GGCGACT ATG AAG GCA GCG ACA GAT CAA GAA TTG AGA AAG CTT    173
             ***  **

      I  V  L  Y  N  N  V  M  E  V  M  E  H  D  A  A  K
 174  ATC GTC TTG TAC AAT AAT GTT ATG GAA GTG ATG GAA CAT GAT GCT GCG AAG   224

      S  M  R  D  D  N  R  A  Y  G  G  F  V  R  A  A  K
 225  TCA ATG CGA GAT GAT AAT CGT GCG TAC GGT GGG TTT GTT CGC GCT GCT AAA   275

      G  K  I  Q  E  L  I  T  E  R  L  V  R  T  V  W  D
 276  GGG AAA ATT CAA GAG CTT ATA ACA GAA AGG CTA GTA AGG ACG GTT TGG GAT   326

      V  E  M  G  E  N  P  E  R  L  S  I  N  S  K  K  I
 327  GTT GAA ATG GGT GAG AAT CCA GAA AGG CTT TCT ATA AAT TCA AAA AAA ATA   377

      K  I  P  I  L  R  S  Y  V  D  S  I  N  D  E  N  L
 378  AAG ATT CCT ATT CTT CGA TCA TAC GTT GAT TCA ATA AAT GAC GAA AAT TTA   428

      K  K  Y  I  S  S  N  I  L  K  Y  S  Y  G  L  S  V
 429  AAA AAA TAC ATA TCT TCA AAT ATT TTA AAA TAT AGC TAT GGT CTT TCT GTA   479

      D  K  H  V  F  I  D  N  K  F  V  L  G  I  E  C  K
 480  GAT AAG CAT GTT TTT ATT GAT AAC AAA TTT GTT TTA GGA ATT GAG TGT AAG   530

      A  Y  T  E  N  A  M  L  K  R  I  L  V  D  F  Y  L
 531  GCA TAC ACA GAA AAT GCA ATG TTA AAA AGA ATA CTT GTA GAT TTT TAT TTG   581

      L  K  T  K  F  P  K  L  N  C  F  L  F  Q  L  E  S
 582  TTA AAG ACA AAG TTT CCA AAA TTG AAT TGC TTT TTG TTT CAA CTT GAA AGT   632

      Q  L  G  G  D  Y  S  E  C  N  K  F  P  I  G  S  Y
 633  CAA CTT GGT GGT GAT TAT TCG GAA TGT AAT AAG TTT CCG ATA GGT AGT TAT   683

      P  T  R  T  I  M  S  Y  F  K  N  V  D  L  N  I  V
 684  CCT ACA AGG ACT ATA ATG TCA TAT TTC AAA AAT GTA GAT TTG AAT ATC GTG   734

      T  L  L  E  G  E  R  K  V  D  R  P  I  N  K  P  Q
 735  ACA TTG TTA GAG GGT GAA AGG AAA GTT GAT CGC CCA ATA AAT AAG CCA CAG   785

      F  F  K  P  L  K  V  E  H  L  E  V  A  I  G  Y  L
 786  TTT TTT AAG CCG CTA AAA GTT GAG CAT TTA GAA GTA GCT ATT GGG TAT CTA   836

      Q  E  S  L  S  E  I  *     M  N  I  I  D  L  F  A
 837  CAA GAA TCA TTA TCG GAG ATT TAG AGG ATG AAT ATA ATT GAT TTA TTT GCA   887
                     *  ***

      G  C  G  G  F  S  H  G  F  K  M  A  G  Y  N  S  I
 888  GGT TGC GGT GGA TTT AGT CAC GGT TTT AAG ATG GCT GGC TAT AAT TCG ATA   938

      L  A  I  E  K  D  L  W  A  S  Q  T  Y  S  F  N  N
 939  CTA GCC ATC GAA AAA GAT CTT TGG GCA TCG CAA ACA TAT TCG TTT AAC AAT   989

      P  N  V  S  V  I  T  E  D  I  T  T  L  D  P  G  D
 990  CCT AAC GTA AGT GTG ATC ACA GAA GAT ATT ACA ACT TTG GAT CCA GGG GAC  1040

      L  K  I  S  V  S  D  V  D  G  I  I  G  G  P  P  C
1041  TTG AAA ATT TCT GTC TCG GAT GTT GAT GGG ATT ATT GGT GGG CCT CCA TGT  1091

      Q  G  F  S  L  S  G  N  R  D  Q  K  D  P  R  N  S
1092  CAG GGT TTT TCT CTT TCT GGA AAT AGA GAC CAA AAA GAT CCA AGA AAT AGT  1142

      L  F  V  D  F  V  R  F  V  K  F  F  S  P  K  F  F
1143  CTT TTT GTT GAT TTT GTT CGA TTT GTG AAA TTC TTT TCT CCA AAG TTT TTT  1193

      V  M  E  N  V  L  G  I  L  S  M  K  T  K  S  R  Q
1194  GTT ATG GAA AAC GTA CTT GGC ATC CTA TCA ATG AAG ACA AAG AGT CGT CAG  1244
```

```
      Y   V   K   D   I   I   A   E   E   F   S   N   V   G   Y   K   V
1245  TAT GTT AAA GAT ATT ATT GCT GAA GAA TTT TCT AAT GTC GGA TAT AAA GTA  1295

      C   V   I   I   L   N   A   C   D   Y   G   V   P   Q   S   R   Q
1296  TGT  GTT ATT ATC TTA AAC GCT TGT GAT TAT GGG GTT CCT CAA TCT CGT CAA 1346

      R   V   F   F   I   G   L   K   S   D   R   P   L   N   Q   Q   I
1347  CGT GTT TTT TTT ATT GGT TTA AAA TCA GAC AGG CCT CTT AAT CAG CAA ATT  1397

      L   T   P   P   S   K   V   I   E   S   E   Y   T   S   L   E   E
1398  TTG ACC CCT CCA TCT AAA GTT ATA GAG TCA GAA TAC ACT TCT CTT GAA GAG  1448

      A   I   S   D   L   P   V   I   E   A   G   E   G   G   E   V   Q
1449  GCG ATC AGT GAT TTG CCA GTG ATA GAG GCG GGT GAA GGA GGG GAG GTG CAG  1499

      D   Y   P   V   A   P   R   N   K   Y   Q   E   N   M   R   K   G
1500  GAT TAT CCC GTT GCT CCT AGG AAC AAG TAT CAA GAA AAT ATG CGG AAG GGA  1550

      S   T   C   V   Y   N   H   V   A   M   R   H   T   Q   R   L   V
1551  TCA ACG TGT GTG TAT AAT CAT GTT GCG ATG CGT CAC ACA CAG CGG CTT GTT  1601

      D   R   F   A   A   I   K   F   G   Q   S   V   K   H   V   S   E
1602  GAT AGA TTC GCC GCA ATT AAA TTC GGC CAA TCT GTT AAG CAT GTC TCT GAA  1652

      E   H   S   Q   R   K   R   G   D   A   N   S   I   S   G   K   V
1653  GAG CAC TCA CAA AGA AAG CGT GGA GAT GCA AAC TCA ATA AGT GGG AAA GTT  1703

      F   S   Q   N   N   M   R   P   Y   P   Y   K   P   C   P   T   V
1704  TTT TCT CAA AAT AAT ATG AGG CCT TAT CCA TAT AAG CCG TGC CCG ACC GTA  1754

      A   A   S   F   Q   S   N   F   I   H   P   F   Y   N   R   N   F
1755  GCT GCA AGT TTT CAG AGT AAT TTT ATT CAT CCA TTT TAC AAT AGA AAT TTT  1805

      T   A   R   E   G   A   R   I   Q   S   F   P   D   T   Y   I   F
1806  ACA GCT CGG GAA GGA GCT AGG ATA CAG TCA TTT CCA GAT ACA TAT ATA TTT  1856

      Q   G   K   R   T   T   M   S   W   E   K   H   L   S   Q   Y   Q
1857  CAA GGT AAA AGA ACT ACA ATG TCC TGG GAA AAG CAT CTT TCT CAA TAT CAG  1907

      Q   I   G   N   A   V   P   P   L   L   A   Q   A   L   A   E   R
1908  CAA ATT GGA AAT GCT GTT CCT CCC TTG CTT GCG CAA GCT CTG GCG GAG AGA  1958

      I   S   W   Y   F   E   N   I   N   L   I   N   D   S   N   V   S
1959  ATT TCT TGG TAT TTT GAA AAT ATA AAT TTG ATA AAC GAC TCA AAT GTA TCG  2009

      I   K   R   M   V   Q   R   S   F   M   S   Q   L   N   L   E   N
2010  ATT AAA AGG ATG GTG CAG CGT TCT TTT ATG TCT CAA TTA AAT TTA GAA AAT  2060

      N   V   N   V   R   Q   D   D   N   Y   D   K   V   H   S   F   *
2061  AAC GTC AAC GTA CGG CAA GAC GAT AAT TAT GAC AAG GTA CAT AGT TTT TAA  2111

      *   S   *   *
2112  TAA AGC TAG TGA TATTT TTTGTTTGTC TAGGTGAAAA GCCTCGCCCA GAAACCGCGC    2168

2169  TTCTGTGCTG GTTGTGGTGC CGCCGTCAAC GCCTTGGCCT GCCATTTGCC ATGCGTCCGT    2228

2229  TGTTCCTGCC ATCGGTCCGT CTGTCCGA                                       2256
```

Figure 3. The nucleotide and amino acid sequences of the 2256bp region containing the *Dde* I endonuclease and methylase genes. The probable -10 and -35 regions are *boxed.* The putative ribosome binding sites are indicated by *asterisks.* The 19 amino acids from the amino terminus of the endonuclease confirmed by protein sequencing are *underlined.*

first 54 bp are necessary for making an active methylase; however it is possible that the 909 deletion is removing part of a promoter sequence and not the structural gene itself.

Upstream of the *r* gene, starting at nucleotide 69 is a sequence highly homologous with the consensus *E. coli* promoter (36). This may in fact be the *r* gene promoter (Figure 3). There is no similar sequence upstream of the *m* gene; considering the close pairing and like orientation of the two genes it is possible that they function as an operon in *D. desulfuricans*. The lack of a strong promoter upstream of the *m* gene may explain its strong dependence on orientation in pBR322 for level of expression (see below). The putative translational start sites of both genes are associated with reasonably good Shine-Dalgarno sequences (37)(Figure 3).

The base composition within the coding region of the genes is 37% G+C, which differs substantially from the 57% G+C content reported for complete genomic *D. desulfuricans* Norway DNA (38).

There are no *Dde*I sites (5'CTNAG3') found contiguous to or within the coding regions for the *Dde*I *r* and *m* genes.

r Gene Expression in pDdeM2.0

Sequence analysis of the *Dde*I *m* gene shows that the termination codon at position 2109 is 102 nucleotides downstream from a *Cla*I site. If the *Dde*I gene boundaries have been correctly assigned, pDdeM1.6, the deletion plasmid containing the "hyperactive" *m* gene, actually encodes a truncated form of the protein (M.*Dde*I'). In this construct the last 33 amino acids (Figure 3) of the *Dde*I methylase (M.*Dde*I), which are highly charged, are replaced by 6 amino acids(MISCQT) encoded by pBR322. We wanted to compare clones containing the truncated and complete forms of the *m* gene, therefore a subclone containing the full open reading frame was constructed (Figure 1).

Clones containing the 3.0kb *Hind*III fragment (pDdeM3.0a and b) in either orientation relative to pBR322 were previously reported to be unable to support the active *r* gene, presumably due to their low levels of methylase activity (1). To determine whether pDdeM2.0 containing the complete *m* gene but lacking an additional 1.0kb of *D. desulfuricans* DNA could support the endonuclease, pDdeR2.3 (containing the complete *r* gene) was introduced into cells containing either pDdeM2.0 or pDdeM1.6. Transformants were tested for both *Dde*I endonuclease activity and for their plasmid composition. Again, only clones containing pDdeM1.6 have *Dde*I endonuclease activity. Cells containing pDdeM2.0 initially produce R.*Dde*I at a very low level. However upon subsequent growth the *r* gene is inactivated by the cell. When the plasmids from the transformants containing pDdeM2.0 are analyzed, various additional fragments due to rearrangement of the pDdeR2.3 plasmid are found.

Effect of Orientation and Deletion on M.*Dde*I Activity

In cloning the *Dde*I system it was necessary to separate the *m* gene from the *r*

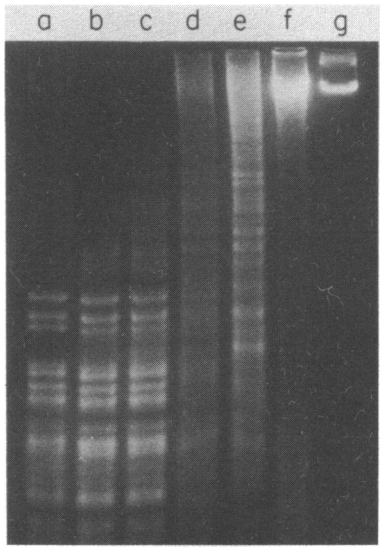


Figure 4. The effect of orientation and deletion on *Dde* I methylase activity. Original *Dde* I methylase clones (pDdeM3.0a and b) and deleted derivatives were tested for *in vivo* methylation of λ_{vir}. λ_{vir} was grown on clones containing each methylase construct. Its DNA was isolated and digested with R.*Dde* I, and resolved on a 1.5% agarose gel. (a) unmodified λ_{vir}; (b) λ_{vir} on pDdeM3.0b; (c) λ_{vir} on pΔDdeM2.1; (d) λ_{vir} on pDdeM3.0a; (e) λ_{vir} on pDdeM2.0; (f) λ_{vir} on pDdeM1.6; and (g) undigested λ_{vir}.

gene, and thus from the promoter like sequence upstream of the *r* gene. If this sequence is in fact the promoter for the complete system, the *m* gene in these methylase clones may be dependent upon pBR322 regions for expression. It was therefore necessary to determine to what extent orientation and flanking sequences were affecting methylase expression.

A series of *Bal*-31 deletions from both the *Eco*RV and *Eco*RI sites of pDdeM3.0a and b was constructed. Levels of methylase expression were assessed both by challenging the deleted plasmid with R.*Dde*I, and by assaying λ DNA that had been grown on the various clones (Figure 4).

Analysis of a series of small *Bal*-31 deletions that eliminate the *Eco*RI and *Hind*III sites shows that removing the Tet promoter has no appreciable effect on methylase activity (data not shown). Deletions in the *r* gene upstream from the start of the *m* gene has a small but consistently positive effect on methylase expression in both orientations. Deletions in the region distal to the *m* gene also increase methylase activity (compare 4: b and c, and 4: d and e). However these effects are minor compared to the difference in activity due to orientation: the "a" orientation is always higher than the "b" orientation

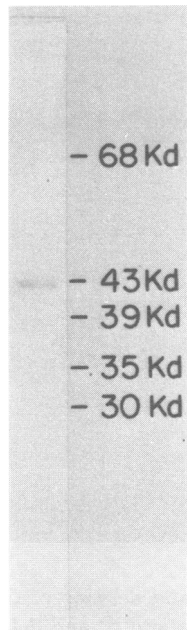


Figure 5. A 12.5% SDS-PAGE of purified M.*Dde* I'. The active purified methylase protein was concentrated on a Vydac C4 214TP54 300A pore reverse phase column, developed with a linear gradient of 5% to 100% acetonitrile in 0.1% TFA with detection at 214 nm. The SDS-PAGE was electroblotted onto Immobilon PVDF transfer membrane (Millipore) and stained with Coomassie blue R250. Molecular weight standards are marked on the side, and indicate that the truncated *Dde* I methylase protein runs as a single band at approximately 43,000d.

(compare 4: b,c vs. d,e). It is possible that the increase in activity is due to the indirect effect of elimination of inhibitory secondary structures in the *D. desulfuricans* DNA and/or decreasing the distance to promoter sequences in pBR322.

Methylase activity high enough to support endonuclease activity and fully protect λ phage is only found in cells with pDdeM1.6, containing the truncated *m* gene. Even when the 2.0kb *Hind*III-*Sac*I fragment containing the complete *m* gene is moved into pUC18, a higher copy plasmid containing the lac promoter, M.*Dde*I activity is still lower than that of cells containing pDdeM1.6 and cannot support R.*Dde*I activity.

It is possible that the difference in activities is due in part to properties of the methylase protein as well as levels of expression. Therefore the complete and truncated methylase proteins were isolated and their activities compared *in vitro*.

Comparison of the Complete (M.*Dde*I) and Truncated (M.*Dde*I') Forms of the Methylase

*Dde*I methylase was purified from clones containing pDdeM1.6 and pDdeM2.0 re-

spectively and compared. The construct carrying the intact methylase appears to yield less activity in cell extracts, and that activity is more labile upon purification and storage than is the activity made by the construct carrying the truncated methylase. The truncated methylase, M.*DdeI'* (from cells containing pDdeM1.6), can be purified to approximately 90% homogeneity by the method outlined in Table 1. When run on SDS-PAGE, the final product gives a major band at approximately 43,000d (Figure 5); this value is consistent with that estimated from the DNA sequence, giving us confidence in the sequence data. This activity is somewhat labile; although the amount of methylase activity found in the crude extracts is high, the yield (1%) and specific activity (8,000 U/mg) of the purified methylase is comparatively low. Purified M.*DdeI'* appears to rapidly lose activity upon storage. After one month storage at -20°C, greater than 80% of the activity is lost.

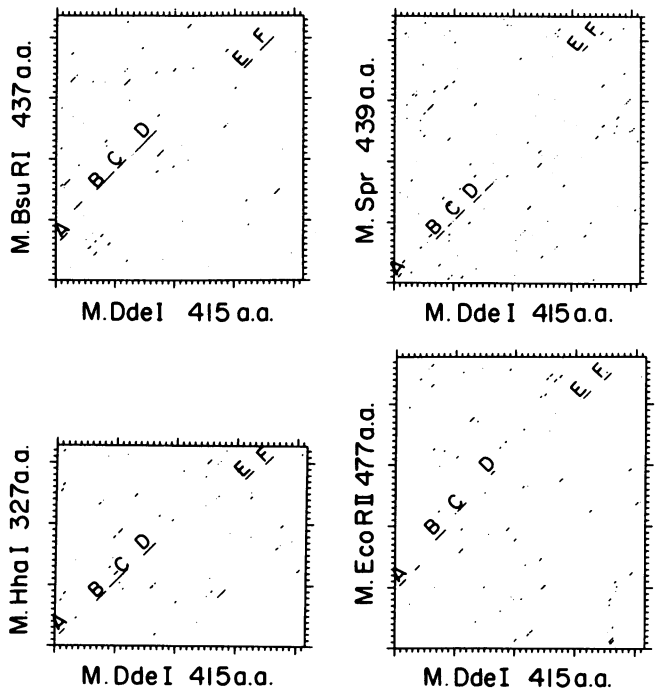
The situation is worse with the intact methylase (M.*DdeI*) from cells containing pDdeM2.0; attempts to purify this protein to homogeneity have been unsuccessful. When starting with an amount of cells similar to that used in purifying M.*DdeI'*, all methylase activity is lost by the second step. When starting with twenty times more cells, there is no recovery of M.*DdeI* activity after the third column. The problem first manifests itself in the crude extracts. It is impossible to get complete protection of one μg of linearized pBR322 with M.*DdeI*, even when twenty times more extract is used than is needed for complete protection with M.*DdeI'*.

The question has arisen as to whether the M.*DdeI'*, being a truncated protein with high activity, has in fact lost some degree of specificity and is modifying sites other than canonical *DdeI* sequences (ie. $5'^{me}$CTNAG 3'). We tested purified M.*DdeI* and M.*DdeI'* for specificity in two different ways. First, the methylase reaction was standardized using several DNA methylases (ie. M.*PstI*, M.*TaqI*, M.*Eco*RI and M.*AluI*). Under saturating conditions, one finds approximately 1.6 X 10^{-7} μmoles of methyl groups incorporated per site for one μg pBR322. Similarly, using M.*DdeI'*, one finds approximately 1.5 X 10^{-7} μmoles/site incorporated. However, with M.*DdeI* only 0.9 X 10^{-7} μmoles/site or 60% incorporation is obtained. Second, pBR322, which had been digested to completion with R.*DdeI* is used as a substrate for methylation by M.*DdeI* and M.*DdeI'*, in reactions using [methyl-^{3}H] S-adenosyl-L-methionine. Neither methylase can incorporate any significant amount of [^{3}H] methyl groups under standard reaction conditions. Therefore it seems unlikely that M.*DdeI'* has lost specificity. It is more likely that M.*DdeI'* is as specific as M.*DdeI* but is also more active. The difference in methylation level *in vitro* reflects the differences seen in the *in vivo* λ protection assays (Figure 4).

Homologous Regions

Using a combination of the dot matrix program from UW (29) and the Align pro-

A



B

	A		B		C	
M.*Dde* I	MNIIDLFAGCGGFSHGFKMAG	1	DGIIGGPPCQGFSLSGNR	68	PKFFVMENVLG	107
M.*Bsu* RI	INVLSLFSGCGGLDLGFELAG	60	NLILGGFPCPGFSEAGPR	150	PEIFVAENVKG	189
M.*Spr*	LRVMSLFSGIGAFEA|GYELVG	4	DLLVGGDPCQSFSVAGHR	70	PKFFVFENVKG	110
M.*Hha* I	LRFIDLFAGLGGFRLALESCG	12	DILCAGFPCQAFSISGKQ	73	PKVVFMENVKN	113
M.*Eco* RII	FRFIDLFAGIGGIRKGFETIG	96	DVLLAGFPCQPFSLAGVS	178	PAIFVLENVKN	217

	D		E		F	
M.*Dde* I	GYKVCVIILNACDYGVPQSR	141	EGARIQSFPDTYIF	318	YQQIGNAVPPLL	347
M.*Bsu* RI	GYRVQFKLLNARDYGVPQLR	220	EIARVQTFPDWFQF	353	YKQIGNAVPVLL	384
M.*Spr*	GYRIDLELLNSKFFNVPQNR	141	ECFRLQAFDDEDFE	384	YKQAGNSITVTV	410
M.*Hha* I	DYSFHAKVLNALDYGIPQKR	144	ECARVMGYPDSYKV	278	YKQFGNSVVINV	299
M.*Eco* RII	GYEVAD|KVIDGKHFLPQHR	258	ECARLMGFEKVDGR	415	YRQFGNSVVVPV	440

Figure 6. Comparison of amino acid sequence of M.*Dde* I with M.*Bsu* RI, M.*Spr*, M.*Hha* I and M.*Eco* RII. (A): Dot matrix evaluation of methylase sequences by compare/dotplot(29). *Diagonal lines* indicate regions of strong homology. Six consistent regions of strong homology (A-F) are labeled and the corresponding sequences are given in table below (B): Amino acid sequences of homologous regions are shown. *Vertical line* in region A of M.*Spr* indicates deletion of amino acids. M.*Bsu* RI is very similar to M.*Bsp* RI and the reported phage methylases *Spr*, ϕ3T and ρ11s are very similar, therefore only one of each group was shown. [Note: our regions E and F are reverse of that reported in (40)].

gram (31), the amino acid sequences of the *Dde*I proteins were compared with other known endonuclease and methylase protein sequences. Comparing R.*Dde*I and M.*Dde*I reveals no significant homology; similarly, comparing R.*Dde*I to other sequenced endonucleases yields no homologies. This is true with other restriction systems thus far characterized (9,34,35,39). Furthermore M.*Dde*I, a DNA cytosine methylase (1), shows no significant homology to any adenine methylase reported to date. This is once again in agreement with what has been previously reported (40).

Regions of homology have been reported among all the prokaryotic DNA cytosine methylases analyzed to date. Once again the *Dde*I methylase is no exception. M.*Dde*I was compared to the published sequences of other cytosine methylases (ie. M.*Bsp*RI, 41; M.*Bsu*RI, 9; M.*Spr*, 42, 43; M.*Hha*I, 5; M.*Eco*RII, 40; ρ11s, 44 and ϕ3T, 45). A representative sampling of these is shown in Figure 6. Homology is found in each set of comparisons. A slightly higher degree of homology is found between M.*Dde*I and M.*Hae*III (B. Slatko, pers. comm.; data not shown). Among the cytosine methylases, there are six regions (A-F) with varying degrees of homology (Figure 6). Four of these regions (A-D) are found near the amino-terminal end of the methylase while the other two homologous regions (E and F) are nearer the carboxy-terminus. In all cases there is a large region of non-homology found between regions D and E. There is evidence in the *Spr* system (42) and, by inference, in the ϕ3T and ρ11s systems (44) that the sequence recognition resides in this non-homologous region, between D and E. This may be the case in other and perhaps all of the related cytosine methylases.

Two additional factors seen from the homologies should be mentioned. There is strong homology at the amino terminus within the A region between M.*Dde*I and many of the other methylases (Figure 6). Since this is one of the most conserved regions among these cytosine methylases, it supports our assignment of the ATG at position 862 to be the start of the *Dde*I methylase. Second, no homology has been found in the region from amino acid 382 to 415 at the carboxy-terminus of M.*Dde*I and any other methylase. This corresponds to the region that has been deleted to form M.*Dde*I'. A similar situation is found in M.*Eco*RII which has a non-conserved region at its amino-terminus. When the first 34 a.a. are deleted, the protein still maintains methylase activity (40).

DISCUSSION

*Dde*I is the only restriction system found to date which has necessitated the cloning of the methylase and endonuclease genes separately. The problem seems to lie in expressing enough M.*Dde*I in *E. coli* to protect against action of R.*Dde*I.

The relatively low level of methylase activity in the cells may be due to poor

transcription. There are no obvious strong promoter sequences associated with the start of the methylase gene. It is possible that the message made is rapidly degraded in *E. coli*. Certainly experiments should be done to analyze M.*Dde*I transcripts.

Another possiblity is that methylase expression is impeded by the presence of the *r* and *m* genes in *cis* on the same plasmid; attempts to reconstruct the complete endonuclease gene on pDdeM1.6 by adding various restriction fragments containing the upstream sequence are always unsuccessful.

Alternatively, the major difficulty may be post-translational. The M.*Dde*I methylase may be subject to a high level of proteolytic degradation in *E. coli*. The difference in activity between M.*Dde*I and M.*Dde*I' may be primarily due to rates of turnover of the two proteins. It therefore may be possible to find protease deficient strains of *E. coli* that will express even the complete form of the methylase gene at sufficiently high levels to support the endonuclease.

Of course, a major unaddressed area is the purification and characterization of M.*Dde*I from *Desulfovibrio*. It would be very interesting to compare it in size and activity to that of the methylase made by *E. coli* clones. Unfortunately these experiments are not easily done since *D. desulfuricans* cells are very difficult to grow and protein yields are low.

It should be added that the *Dde*I clones are not completely viable; upon prolonged storage on Petri dishes or frozen at -70°C, the cells die. Therefore, we are continuing efforts to identify and solve problems inherent in the *Dde*I clones. At the same time, we are using the information gained to clone other refractory systems.

ACKNOWLEDGMENTS

The authors wish to thank Carla Vaccaro for excellent technical assistance and Dr. Jack Benner for help in using the various computer programs. We also thank Drs. R. Blumenthal, A. Bhagwat, T. Gingeras, J. Benner, E. Raleigh, I. Shildkraut and D. Comb for helpful criticism of the manuscript.

*Present address: Department of Biochemistry, Brandeis University, Waltham, MA 02254, USA

REFERENCES

1. Howard, K.A., Card, C., Benner, J.S., Callahan, H.L., Maunus, R., Silber,K., Wilson, G., and Brooks, J.E.(1986)Nucl. Acids Res. 14, 7939-7951.
2. Yuan, R. (1981) Ann. Rev. Biochem. 50, 285-315.
3. Arber, W. (1974) in Progress in Nucleic Acid Research and Molecular Biology, Cohn, W.E. Ed. Vol. 14, pp.1-37, Academic Press, New York.
4. Smith, H.O. (1979) Science 205, 455-466.

5. Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G. and Wells, R. (1987) J. Biol. Chem. 262, 4770-4777.
6. Walder, R.Y., Langtimm, C.J. and Chatterjee, R. (1983) J. Biol.Chem. 258, 1235-1241.
7. Szomolanyi, E., Kiss, A. and Venetianer, P. (1980) Gene 10, 219-225.
8. Mannarelli, B.M., Balganesh, T.S., Greenberg, B., Springhorn, S.S. and Lacks, S.A. (1985) Proc. Natl. Acad. Sci. USA 82, 4468-4472.
9. Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) Nucl. Acids Res. 13, 6403-6421.
10. Slatko, B., Moran, L., Simcox, T., Benner, J., Jager, T. and Wilson, G., Nucl. Acids Res., submitted.
11. Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, B. (1986) Cell 46, 993-1000.
12. Blumenthal, R.M., Gregory, S.A. and Cooperider, J.S. (1985) J. Bacteriol. 164, 501-509.
13. Raleigh, E.A. and Wilson, G. (1987) Proc. Natl. Acad. Sci. USA 83, 9070-9074.
14. Ross, T.K. and Braymer, H.D. (1987) J. Bacteriol. 169, 1757-1759.
15. Noyer-Weidner, M., Diaz, R. and Reiners, L. (1986) Mol. Gen. Genet. 205, 469-475.
16. Heitman, J. and Model, P. (1987) J. Bacteriol. 169, 3243-3250.
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratories, New York.
18. Yanisch-Perron, C., Viera, J. and Messing, J. (1985) Gene 33, 103- 119.
19. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA 62, 1159-1166.
20. Gingeras, T.R. and Brooks, J.E. (1983) Proc. Natl. Acad. Sci. USA 80, 402-406.
21. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. USA 69, 2110-2114.
22. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
23. Hattori, M. and Sakaki, Y. (1986) Analytical Biochemistry 152, 232- 238.
24. Chen, E.Y.and Seeburg, P.H. (1985) DNA 4,165.
25. Haltiner, M., Kempe,T., and Tjian,R. (1985) Nucl. Acids Res. 13, 1015-1025.
26. Olsson,A., Moks,T., Muhlen,M., and Gaal, A.B. (1984) J. Biochem. Biophys. 10, 83-90.
27. Ansorge, W. and Labeit, S. (1984) J. Biochem. Biophys. 10, 237-243.
28. Williams,S.A., Slatko,B.E., Moran,L.S., and DeSimone, S.M. (1986) BioTechniques 4, 138-147.
29. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. 12, 387-395.
30. Keller, C., Corcoran, N. and Roberts, R.E. (1984) Nucl. Acids Res. 12, 379-386.
31. Colowick, S.P. and Kaplan, N.O. (1983) in Methods in Enzymology, Hirs, C.H.W., Timasheff, S.N., ed., Vol. 91 pp. 524-540, Academic Press, New York.
32. Theriault, G., Roy, P.H., Howard, K.A., Benner,J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) Nucl. Acids Res. 13, 8441-8461.
33. Bradford, M.M. (1976) Analyt. Biochem. 72, 248-254.
34. Newman, A.K., Rubin, R.A., Kim, S.H. and Modrich, P. (1981) J. Biol. Chem. 256, 2131-2137.

35. Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J.M. (1981) J. Biol. Chem. 256, 2143-2153.
36. Gouy, M. and Gautier, C. (1982) Nucl. Acids Res. 10, 7055-7074.
37. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.
38. Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. (1985) Nucl. Acids Res. 13, 1399-1412.
39. Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) J. Biol. Chem. 259, 8015-8026.
40. Som, S., Bhagwat, A.S. and Friedman, S. (1987) Nucl. Acids Res. 15, 313-331.
41. Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) J. Mol. Biol. 170, 597-610.
42. Buhk, H.J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Gunthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) Gene 29, 51-61.
43. Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) Nucl. Acids Res. 12, 9039-9049.
44. Behrens, B., Noyer-Weidner, N. Pawlek, B., Lauster, R., Balganesh, T.S. and Trautner, T.A. (1987) EMBO J 6, 1137-1142.
45. Tran-Becke, A., Behrens, B., Noyer-Weidner, N. and Trautner, T.A. (1986) Gene 42, 89-96.