Volume 15 Number 23 1987 Nucleic Acids Research

Cloning, sequencing and expression of the *Taq* I restriction–modification system

Barton E.Slatko*, Jack S.Benner, Tineka Jager-Quinton, Laurie S.Moran, Timothy G.Simcox, Elizabeth M.Van Cott and Geoffrey G.Wilson

New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915, USA

Received August 28, 1987; Revised and Accepted November 4, 1987 Accession no. Y00499

ABSTRACT

The *Taq* I modification and restriction genes (recognition sequence TCGA) have been cloned in *E.coli* and their DNA sequences have been determined. Both proteins were characterized and the N-terminal sequence of the endonuclease was determined. The genes have the same transcriptional orientation with the methylase gene 5' to the endonuclease gene. The methylase gene is 1089 bp in length (363 amino acids, 40,576 daltons); the endonuclease gene is 702 bp in length (234 amino acids, 27,523 daltons); they are separated by 132 bp. Both methylase and endonuclease activity can be detected in cell extracts. The clones fully modify the vector and chromosomal DNA but they fail to restrict infecting phage. Clones carrying only the restriction gene are viable even in the absence of modification. The restriction gene contains 7 *Taq* I sites; the modification gene contains none. This asymmetric distribution of sites could be important in the regulation of the expression of the endonuclease gene.

INTRODUCTION

Over 100 different Type II restriction endonucleases have been identified among microorganisms (1,2). Each endonuclease recognizes a distinct DNA sequence or set of sequences, and each occurs, for the most part, together with a modification methylase of identical sequence specificity. Restriction endonucleases cleave DNA at, or close to, their recognition sequences. Modification methylases alter the sequences by methylation of internal adenine or cytosine residues; once methylated, the sequences are no longer susceptible to cleavage by the endonuclease. The function of restriction-modification systems is thought to be protective: the endonuclease protects the cell from infection, by digesting unmodified, foreign DNA; the methylase protects the cell from auto-digestion, by modifying its DNA.

The bacterium *Thermus aquaticus* YT1 possesses at least two restriction-modification systems (3,4). The *Taq* I system recognizes double stranded DNA at the palindromic sequence TCGA. The *Taq* I endonuclease cleaves both strands between the T and C residues leaving a 2-base 5' overhang (3); the *Taq* I methylase modifies the adenine residue on each strand at the N6 position to form TCG^{me}A (5).

The genes for 27 Type II restriction-modification systems have now been cloned into *E.coli* (6,7,8,9,10,11,12,13,14,15,16,17, Wilson *et al*, in preparation) and in each of these cases, the genes have been found to be closely linked. This linkage presumably reflects their functional symbiosis and probable co-evolution. The linkage between restriction and modification genes facilitates their cloning because it enables both genes to be isolated together on a single, or on a small number of adjacent, restriction fragment(s).

The genes for the *Taq* I restriction-modification system have been cloned into *E. coli*. The isolation and properties of the clones and the sequence of the genes is presented in this report.

MATERIALS AND METHODS

Reagents

Enzymes, dephosphorylated pBR322, DNA sequencing reagents and primers were from New England Biolabs, Inc. and were used according to manufacturers instructions.

Purification of *T. aquaticus* and plasmid DNAs

Thermus aquaticus YT1, originally supplied by J.I.Harris, was grown to saturation in Castenholz TYE Medium (18) at 70°C, harvested and stored at -70°C. Chromosomal DNA was prepared as previously described (19). Plasmid DNA was extracted from cells by the lysozyme/Triton X-100 lysis procedure and purified by CsCl/EtdBr density gradient ultracentrifugation (20). Plasmid mini-preparations were made by a modification of the alkaline-SDS procedure (21).

Transformation

E.coli strains RR1, MM294 and K802 were made competent by a modification of the calcium, heat-shock method (22,23). Transformation mixes were plated on Luria-agar plates (24) containing antibiotics (ampicillin, 100 µg/ml; kanamycin, 100 µg/ml; tetracycline, 20 µg/ml; chloramphenicol, 200 µg/ml) to select for transformants.

Endonuclease and methylase assays

Overnight cultures were centrifuged and concentrated to an OD_{590} of 75 by resuspension in lysis buffer (0.01M Tris 0.001M EDTA, pH 7.5, 10 mM β-mercaptoethanol (BME), 1 mg/ml lysozyme). After two hrs on ice, the suspension was frozen at -20°C, thawed on ice, and mixed with an equal volume of lysis buffer without lysozyme, but containing 0.01% Triton X-100. Aliquots were microcentrifuged and the supernatant was removed and either assayed immediately or partially purified by heating to 65°C for 15 min to co-precipitate the remaining DNA and thermolabile protein.

For restriction endonuclease assays, 7.5 µl of extract was mixed with 150 µl of digestion buffer (10mM Tris pH 8.4, 10mM $MgCl_2$, 100 mM NaCl, 10mM BME) containing 50 µg/ml lambda DNA. The solution was serially diluted in

3.3-fold steps to produce a titration range of 1 to 0.001 µl of extract per µg DNA. After incubation at 65°C for 1 hr, an aliquot from each solution was then electrophoresed in 1% agarose. A unit of *Taq* I endonuclease activity is defined as that required to digest 1 µg of lambda DNA to completion in 1 hour at 65°C.

For modification methylase assays, 7.5 µl of extract was mixed with 150 µl of methylation buffer (50mM Tris, 10mM EDTA pH 8.0, 5mM BME, 0.1 mM S-adenosylmethionine) containing 100 µg/ml lambda DNA. Serial dilutions and incubations were performed as described above, then the volumes of each were doubled and buffer conditions were converted to those suitable for digestion. Ten units of *Taq* I per µg DNA were added to each tube and the solutions were again incubated at 65°C for 1 hr. Aliquots were then electrophoresed to determine the extent of protection from digestion that the first incubation had provided. A unit of *Taq* I methylase activity is defined as that required to completely protect 1 µg of lambda DNA in 1 hour at 65°C.

The assays were found to be improved by heat-treating the extract (65°C for 15 min), as mentioned above. During this step, a thick coagulate of protein and nucleic acid forms and is removed by brief microcentrifugation. Because the *Taq* I enzymes are thermostable, they were expected to be unaffected by the heat treatment and to remain in solution. Control experiments demonstrated that at least for the endonuclease, the titer did not diminish during heat-treatment throughout a wide range of extract concentrations. Heat-treatment of large volumes of extract intended for enzyme purification, however, abolishes some chromatographic properties of the endonuclease, as the enzyme appears to loose affinity for resins on which it normally separates, making its purification extremely difficult.

DNA Sequencing.

Sanger dideoxy DNA sequencing (25) was performed by the method of Williams et al (26), or by a modification of a "collapsed plasmid" protocol (27,28,29). All reactions utilized [α-^{35}S]-dATP (600 - 1200 Ci/mM, New England Nuclear) and were electrophoresed on 6-8% standard, "wedge" or buffer-gradient sequencing gels (30,31,32). Both DNA strands in the region encompassing the *Taq* I endonuclease and methylase genes were completely sequenced by using templates made from *Sau3A* I and *Msp* I subclones from gel-purified fragments containing the *Taq* I methylase and endonuclease genes ligated into the *BamH* I and *Acc* I polylinker sites of M13mp18 and M13mp19 (23), by double digests of the gel-purified fragments "forced-cloned" into the two M13mp vectors, and by the use of synthetic oligonucleotide primers synthesized on a Biosearch 8600 automated synthesizer. Ligations were transformed into *E. coli* strain ER1451 (34). Data was entered and analyzed utilizing a Grafbar Digitizer (SAC

Corporation) and computer programs from Cold Spring Harbor Laboratories (35) and from The University of Wisconsin Computer Group (36).

Protein purification and sequencing

Two hundred grams of frozen *E. coli* RR1 (pSW149RM-3A) cells, were resuspended in 400 ml of ice-cold buffer P (0.1M NaCl, 10 mM KPO_4, 5 mM BME, 0.1 mM EDTA, pH 6.9) and the solution was brought to a final concentration of 200 µg/ml with lysozyme. The solution was maintained at 4°C for 1 hr and then subjected to several 1 min periods of sonication. After cellular debris was removed by centrifugation, the supernatant was removed, adjusted to pH 6.9, and applied to a P11 phosphocellulose column (Whatman) (5.0 x 8.0 cm). After washing with two volumes of buffer P, the column was developed with a linear gradient of 0.1M to 1.0M NaCl in buffer P and fractions were assayed for endonuclease and methylase activity. The endonuclease eluted at 0.15 M NaCl and the methylase eluted at 0.25 - 0.30 NaCl. The most active fractions of each were pooled and dialyzed against buffer P (pH 7.4).

The endonuclease pool from the P11 column was applied to a Heparin-Sepharose (Pharmacia) column (2.5 x 2.5 cm) which had been equilbrated in buffer P, pH 7.4 and washed with two volumes of the same buffer. The column was developed with a linear gradient of 0.1M to 1.0M NaCl in buffer P (pH 7.4) and fractions were assayed for endonuclease activity. To concentrate the endonuclease, the active fractions were applied to a hydroxylapatite (Calbiochem) column (2.5 x 10.0 cm), which had been equilbrated with buffer P (pH 7.4). The enzyme was step eluted from the column using buffer H (0.1M NaCl, 1.0M, KPO_4, 5 mM BME, 0.1 mM EDTA, pH 7.4), and reapplied to a P11 column (1.5 x 5.0 cm). The column was washed with two volumes of buffer P and developed with a linear gradient of 0.1M to 1.0M NaCl in buffer P. Fractions containing endonuclease activity were pooled and dialyzed against buffer 0.05M KCl, 10mM $KP0_4$, 5mM BME, pH 7.4.

The *Taq* I endonuclease was passed through a Mono S column (Pharmacia), dialyzed against 20mM Tris-HCl (pH 7.5), 50mM KCl, 10mM, BME, absorbed to a Mono Q column (Pharmacia) and eluted with a linear gradient from 50 to 900mM KCl. The endonuclease activity eluted as a single 280nm peak at approximately 250mM KCl. The peak produced a single band of 30,000 daltons when subjected to SDS-gel electrophoresis and stained using Kodavue (Eastman).

The *Taq* I methylase pool from the P11 column was applied to a Heparin-Sepharose (Pharmacia) column (2.5 x 2.5 cm) which had been equilbrated in buffer P, pH 7.4 and washed with two volumes of the same buffer. The column was developed with a linear gradient of 0.1M to 1.0M NaCl in buffer P (pH 7.4) and the most active fractions were pooled. The methylase was loaded on a Mono S column (Pharmacia) and eluted at a flow rate of 0.5 ml/min with a linear gradient from 50 to 900mM KCl in 20mM KPO_4, 10mM BME, (pH 6.9). The *Taq* I methylase activity eluted as a single UV absorbing peak (280 nm) at

approximately 250mM KCl. The methylase fractions were then applied to a Sephadex G-100 (Pharmacia) column (2.5 x 80.0 cm), which had been equilbrated with buffer P (pH 7.4), and the column effluent was monitored at 280 nm. The methylase eluted from the column at an elution volume corresponding to a molecular weight of 40,000 daltons and the active fractions were pooled.

The *Taq* I methylase was concentrated by loading on a Mono S column (Pharmacia) and eluted as before. The *Taq* I methylase activity eluted as a single UV absorbing peak (280 nm) of approximately 20 ug of protein at approximately 250mM KCl. The peak was concentrated using a Centricon filter (Amicon) and subjected to SDS-gel electrophoresis. Staining using Kodavue (Eastman) revealed multiple protein bands. Due to the small amount of protein containing methylase no further purification was attemped.

Before sequencing, the endonuclease sample was subjected to a final chromatography on a Vydac C4 214TP54 (5um, 4.6 X 300mm) 300 A pore reverse phase column, developed with a linear gradient of 5% to 100% acetonitrile in 0.1% trifluroacetic acid with detection at 214 nm. Individual peaks were manually collected and lyophilized.

The sequential degradation of proteins was performed with an Applied Biosystems model 470A gas-phase sequenator using no-vacuum chemistry (37). The first twenty-seven phenylthiohydantoins were unambiguously identified by high- performance liquid chromatography on an IBM Cyano (5um, 4.5 X 250mm) column with slight gradient modifications from those previously described (38). A Pharmacia fast protein liquid chromatograph (FPLC) was used for Pharmacia Mono Q, Mono S and PolyCAT A columns. A Waters Associates Liquid Chromatograph was used for C4 and Cyano reverse phase chromatography.

RESULTS

Isolation of Taq I Restriction and Modification Clones

Plasmid recombinants carrying the *Taq* I modification gene were selectively isolated from libraries of *T. aquaticus* DNA by the procedure proposed by Mann et al (6). The libraries were prepared by ligating restriction fragments of *T. aquaticus* DNA into pBR322 and propagating the plasmids in *E. coli* to allow self-modification of methylase clones to occur. The selections were performed by digesting the libraries with *Taq* I endonuclease to destroy unprotected molecules and transforming the mixtures back into *E. coli* to recover survivors.

Modified clones were found at a high frequency among the survivors of libraries, prepared with *Pst* I and *Bam*H I. Ninety percent of the survivors from the *Pst* I library were found to be completely resistant to *Taq* I digestion, and to carry a single 3.5 kb *Pst* I fragment in common. Extracts from three clones were prepared and found to contain 2.5 X 10^3 units/ml of extract of *Taq* I methylase, but no detectable endonuclease. One of the

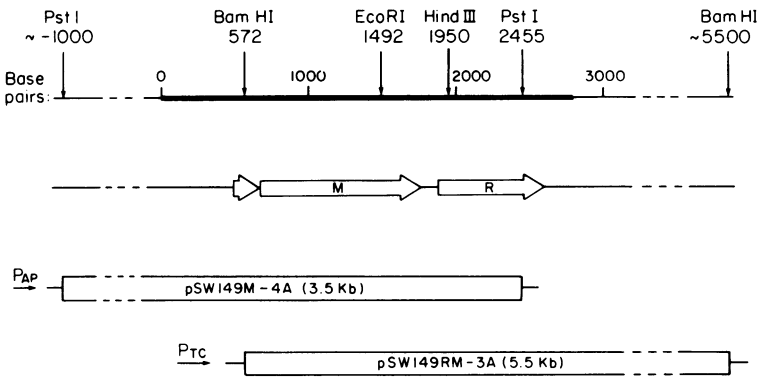

FIGURE 1: Organization of the *Taq* I R and M genes. The locations of the *Pst* I and *Bam*H I fragments cloned into pBR322 to make the Taq I M+ plasmid pSW149M-4A and the *Taq* I R^+ M^+ plasmid pSW149RM-3A are shown. The heavy section of the bp scale indicates the extent of the DNA that was sequenced.

clones, pSW149M-4A, was retained for further study. Fifty percent of the survivors from the *Bam*H I library were found to be completely resistant to *Taq* I digestion and to carry a 5.5 kb *Bam*H I fragment in common. Extracts from these clones were found to contain 2.5 X 10^3 units/ml of extract of *Taq* I methylase and 5 X 10^3 units/ml of extract of *Taq* I endonuclease. One of the clones, pSW149RM-3A, was retained for further study.

Restriction mapping of the *Pst* I and *Bam*H I clones indicated that the fragments overlapped by approximately 2 kb (Figure 1). DNA sequencing confirmed that the overlap region contains the complete methylase gene and most of the endonuclease gene as well, with the unique portion of the *Bam*H I fragment encoding only a short section at the C-terminus of the endonuclease.

Orientation-dependent expression of the methylase gene

All of the *Pst* I clones were found to carry the fragment in one orientation, that in which the methylase gene lay in the same direction as the Ap^R gene of the vector. Similarly, the *Bam*H I clones were also all found to carry the fragment in one orientation, that in which the gene lay in the same direction as the Tc^R gene of the vector. To test the assumption that the expression of the methylase gene derived from the plasmid promoters, P_{Ap} and P_{Tc}, respectively, the fragment orientations were reversed. In both experiments, several individuals with the parental orientation and several with the reverse-orientation were isolated. The parental-orientation plasmids all displayed complete resistance to *Taq* I digestion and the reverse-orientation plasmids all displayed complete sensitivity, strongly suggesting that transcription of the methylase gene derives from promotors on the vector. Methylase assays of extracts of each

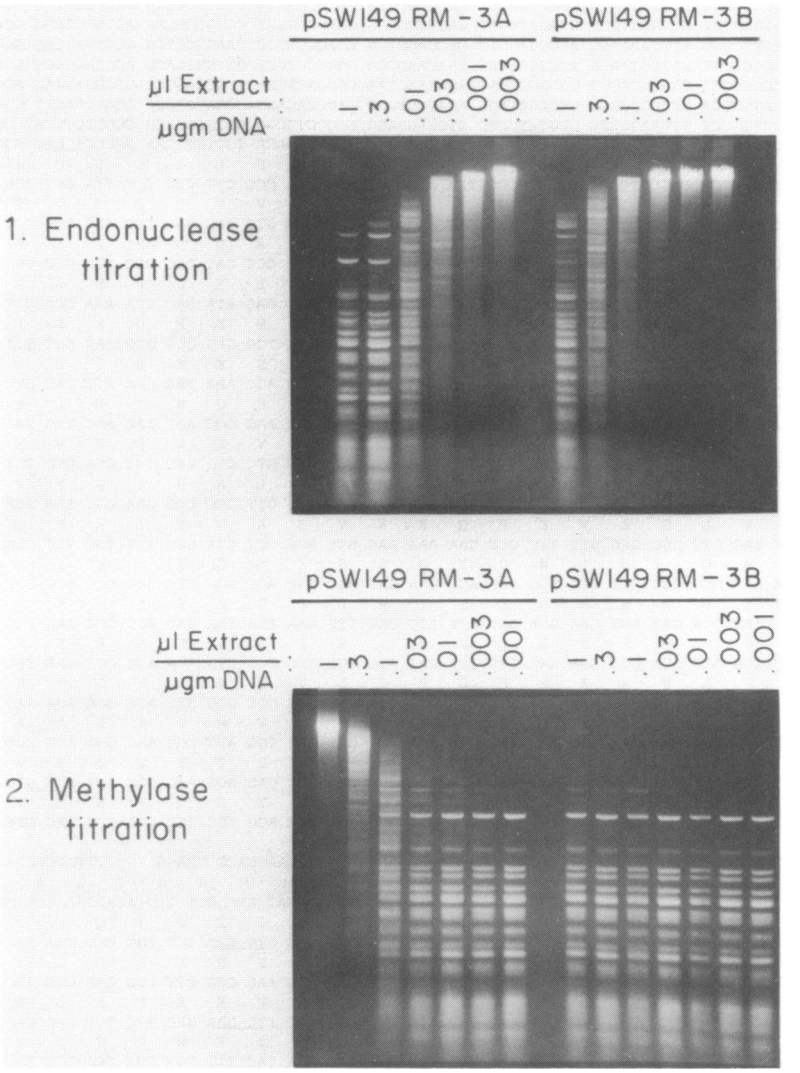

FIGURE 2: *Taq* I endonuclease (1) and methylase (2) assays of cell extracts of *E.coli* RR1 containing cloned *Taq* I R and M genes. The assays were performed as titration series on phage λ DNA. The first extract for each assay was prepared from the parental *Bam*H I clone, pSW149RM-3A, and displays approximately 5 X 10^3 Units of endonuclease and 1 X 10^3 Units of methylase per ml of extract. The second extract, prepared from a reverse-orientation derivative of the *Bam*H I clone, pSW149RM-4B, displays approximately 20% of the parental endonuclease activity and no more than 1% of the parental methylase activity.

```
  1  GGCCTCAAGC TCATCGTACA CCGACGGGGT CTTCTCCATG GACGGGACAT CGCCCCCCTG GACCGCATCG TCCCCCTGGC
 81  CAAGAAGTAC GGGGCGTGGT CTACGTGGAC GACGCCCACG GAAGCGGGGT CCTGGGGGAA AGGGGCGAGG GCACGGTGCA
161  CCACTTCGGC TTCCAGAAGG ACCCGACGTG GTCCAGGTGG CCACCCTTTC CAAGGCCTGG GCGTGATGGC GGGTACGCCG
241  CCGGGGCCAT GAGCTCAAGG AGCTCCTCAT CAACAAGCCA GGCCCCTCCT CTTCTCCACC ACCCACCCCC CGGCGGTGGT
321  GGGGGCCCTC CTGGGCGCCC TGGAGCTCAT AGAGAAGGAG CCCGATAGGA TCGCCAGGCT CTGGGAGAAC ACCCGCTACT
401  TCAAGGCCGA GCTGGCCCGG CTGGGCTACG ACACCCTGGG GAGCCAGACC CCCATCACCC CGGTCCACTT CGGCGAGGCC
481  CCCCTGGCCT TTGAGGCGAG CCGCATGCTC CTGGAGGAGG GGGTCTTCGC CGTGGGCATC GGCTTCCCCA CCGTGCCCCG
561  GGGGAAGGCC AGGATCCGCA ACATCGTCAC CGCCGCCCAC ACGGTGGAGA TGCTGGACAA GGCCCTCGAG GCCTACGAGA
                              Methylase:  M  G  L  P  P  L  L  S  L  P  S  N  A
641  AGGTGGGCCG CAGCGTCGGT ATAATCCGC TAA ATG GGC CTG CCA CCC CTT CTG TCC TTA CCT TCC AAC GCC
       A  P  R  S  L  G  R  V  E  T  P  P  E  V  V  D  F  M  V  S  L  A
712  GCC CCC AGG AGC CTG GGC CGG GTG GAG ACC CCC CCG GAG GTG GTG GAC TTC ATG GTC TCC CTG GCC
       E  A  P  R  G  G  R  V  L  E  P  A  C  A  H  G  P  F  L  R  A  F
778  GAG GCG CCC AGG GGG GGA AGG GTG CTG GAG CCC GCC TGC GCC CAT GGG CCC TTC CTC CGG GCT TTC
       R  E  A  H  G  T  A  Y  R  F  V  G  V  E  I  D  P  K  A  L  D  L
844  CGG GAG GCC CAC GGG ACG GCC TAC CGC TTC GTG GGG GTG GAG ATA GAC CCA AAA GCC CTG GAC CTC
       P  P  W  A  E  G  I  L  A  D  F  L  L  W  E  P  G  E  A  F  D  L
910  CCC CCC TGG GCC GAG GGC ATC CTG GCG GAC TTC CTC CTC TGG GAG CCG GGG GAG GCC TTT GAC CTG
       I  L  G  N  P  P  Y  G  I  V  G  E  A  S  K  Y  P  I  H  V  F  K
976  ATC CTG GGC AAT CCG CCT TAC GGC ATC GTA GGA GAA GCC AGC AAA TAC CCC ATT CAC GTC TTC AAA
       A  V  K  D  L  Y  K  K  A  F  S  T  W  K  G  K  Y  N  L  Y  G  A
1042 GCG GTC AAG GAC CTC TAC AAG AAG GCC TTT TCC ACC TGG AAG GGC AAG TAC AAC TTG TAC GGG GCC
       F  L  E  K  A  V  R  L  L  K  P  G  G  V  L  V  F  V  V  P  A  T
1108 TTT CTT GAA AAG GCC GTT CGC CTT CTT AAG CCT GGT GGG GTC CTC GTC TTT GTA GTC CCG GCC ACC
       W  L  V  L  E  D  F  A  L  L  R  E  F  L  A  R  E  G  K  T  S  V
1174 TGG CTT GTC CTG GAG GAT TTT GCC CTC CTT CGC GAG TTC CTT GCC CGG GAA GGG AAA ACA TCT GTA
       Y  Y  L  G  E  V  F  P  Q  K  K  V  S  A  V  V  I  R  F  Q  K  S
1240 TAC TAC CTT GGC GAG GTT TTC CCG CAA AAA AAG GTT AGC GCT GTA GTG ATT CGC TTC CAG AAG AGC
       G  K  G  L  S  L  W  D  T  Q  E  S  E  S  G  F  T  P  I  L  W  A
1306 GGA AAA GGC CTT TCA CTT TGG GAT ACC CAA GAA AGC GAA AGC GGG TTC ACG CCC ATC CTC TGG GCT
       E  Y  P  H  W  E  G  E  I  I  R  F  E  T  E  E  T  R  K  L  E  I
1372 GAA TAT CCA CAT TGG GAA GGA GAG ATT ATC CGC TTT GAA ACA GAG GAG ACG CGG AAG CTG GAA ATA
       S  G  M  P  L  G  D  L  F  H  I  R  F  A  A  R  S  P  E  F  K  K
1438 TCG GGA ATG CCA CTG GGA GAC CTC TTT CAT ATC CGC TTC GCC GCA AGA AGC CCT GAA TTC AAG AAA
       H  P  A  V  R  K  E  P  G  P  G  L  V  P  V  L  T  G  R  N  L  K
1504 CAT CCA GCA GTG AGA AAG GAA CCG GGG CCA GGT CTT GTG CCT GTG CTC ACA GGA AGA AAT TTA AAG
       P  G  W  V  D  Y  E  K  N  H  S  G  L  W  M  P  K  E  R  A  K  E
1570 CCG GGG TGG GTA GAT TAC GAG AAA AAC CAC TCC GGG CTT TGG ATG CCC AAG GAA AGG GCC AAG GAG
       L  R  D  F  Y  A  T  P  H  L  V  V  A  H  T  K  G  T  R  V  V  A
1636 CTC AGG GAC TTC TAT GCC ACG CCC CAC CTG GTG GTA GCC CAC ACC AAG GGG ACT AGA GTG GTG GCC
       A  W  D  E  R  A  Y  P  G  G  R  S  S  T  S  C  P  R  K  V  *
1702 GCT TGG GAC GAA AGG GCC TAC CCT GGC GGG AGG AGT TCC ACC TCC TGC CCA AGG AAG GTG TGA GAC
       *
1768 TAG ACCCCCGTCC CTGGTGCAGT GGTTAAACTC CGAAGCGATG CAGAAGCACG TCAGGACGCT TTATCGCGAC TTCGTG
                                                  Endonuclease:  M  A  S  T  Q  A  Q  K
1847 CCCCACCTGA CGCTGAGGAT GCTAGAAAGG CTTCCTGTAA GGAGGGAAT ATG GCT TCC ACA CAA GCC CAG AAA
       A  L  E  T  F  E  R  F  L  A  S  L  D  L  E  S  Y  Q  Q  K  Y  R
1921 GCG CTC GAA ACT TTT GAG CGT TTT CTC GCA AGC TTG GAC CTC GAG TCC TAC CAG CAA AAG TAC CGC
       P  I  K  T  V  E  Q  D  L  P  R  E  L  N  P  L  P  D  L  Y  E  H
1987 CCT ATC AAA ACG GTT GAA CAA GAC CTG CCT AGG GAG CTG AAC CCG CTT CCG GAC CTG TAC GAG CAT
       Y  W  K  A  L  E  D  N  P  S  F  L  G  F  E  E  F  F  D  H  W  W
2053 TAT TGG AAA GCG CTT GAG GAT AAC CCT TCC TTC CTG GGC TTC GAA GAG TTC TTT GAC CAC TGG TGG
       E  K  R  L  R  P  L  D  E  F  I  R  K  Y  F  W  G  C  S  Y  A  F
2119 GAA AAG CGC CTA CGG CCC TTG GAC GAG TTC ATA CGC AAA TAC TTT TGG GGA TGC TCC TAC GCG TTT
       V  R  L  G  L  E  A  R  L  Y  R  T  A  V  S  I  W  T  Q  F  H  F
2185 GTT CGC TTG GGC CTC GAG GCT AGG CTG TAC CGA ACA GCC GTT TCC ATC TGG ACT CAG TTT CAC TTC
       C  Y  R  W  N  A  S  C  E  L  P  L  E  A  A  P  E  L  D  A  Q  G
2251 TGC TAC CGC TGG AAC GCC TCC TGC GAG CTT CCT CTA GAA GCT GCC CCA GAA CTC GAC GCC CAA GGG
       I  D  A  L  I  H  T  S  G  S  S  T  G  I  Q  I  K  K  E  T  Y  R
2317 ATA GAC GCG CTG ATT CAT ACA AGC GGG TCC TCA ACA GGA ATC CAG ATC AAA AAG GAA ACT TAC CGT
       S  E  A  K  S  E  N  R  F  L  R  K  Q  R  G  T  A  L  I  E  I  P
2383 TCC GAG GCC AAG AGC GAG AAC CGC TTT TTA AGG AAG CAA AGA GGC ACC GCC CTC ATC GAG ATT CCC
       Y  T  L  Q  T  P  E  E  L  E  E  K  P  T  G  K  S  E  R  R  N  L
2449 TAC ACC CTG CAG ACA CCA GAG GAG CTC GAA GAA AAG CCA ACG GGC AAG AGT GAA CGG AGA AAC CTA
       P  S  M  G  Q  G  C  T  P  F  G  P  S  R  K  R  I  R  H  F  S  G
2515 CCG TCT ATG GGC CAA GGT TGC ACA CCA TTT GGA CCG TCT AGA AAA CGG ATT CGT CAT TTT TCG GGA
       K  L  C  E  K  H  *
2581 AAG TTA TGT GAA AAG CAT TGA GCTT TTTCTCCAGA AAAACGCTCC TACCCTATCT GGGCTCATCC GCTGGGACAG
2656 GGTGGCCCAG GAAGCCCTCA CCGCCCCGTG AGGTAGACAC GAAGCACAAG CCCCACAGCA AAGAGCAGCC CACGGCCACG
2736 TAGACCTCGG GGCGCTTGGG GGGTCCTTGA GACCCCGG
```

of the clones showed that those with the parental orientation synthesized 2.5 X 10^3 units/ml of extract while no methylase could be detected in clones with the reverse orientation (Figure 2). Endonuclease assays of the *Bam*H I clones revealed that they synthesized endonuclease in both orientations: extracts of the parental clones contained 5 X 10^3 units/ml of extract and those of the reverse-orientation clones, about 1 X 10^3 units/ml of extract (Figure 2). The finding that the reverse-orientation clones were unmodified accounted for their absence among the survivors of the libraries: without modification there was not selective protection.

Phage restriction

The parental R^+M^+ *Bam*H I clone, pSW149RM-3A,and a reverse-orientation R^+M^- derivative, pSW149RM-3B, were tested for their ability to plate unmodified stocks of the lambdoid phages 80, 81, 170, and lambda. The phages each plated with an efficiency of 1, relative to their plating on control cells carrying only pBR322. Possession of the *Taq* I endonuclease does not, therefore, enable the cells to restrict phages.

DNA sequence

The sequence of the DNA common to the *Pst* I and *Bam*H I fragments was determined together with approximately 500 preceding nucleotides from the *Pst* I fragment and approximately 300 succeeding nucleotides from the *Bam*H I fragment. Two major open reading frames were identified, oriented in the same direction (Figure 3). The first open reading frame, 1089 bases in length, from coordinates 673 to 1762 was assigned to the methylase gene (363 amino acids, molecular weight= 40,576). The second open reading frame, 702 bases in length, from co-ordinates 1897 to 2599, was assigned to the restriction endonuclease gene (234 amino acids, molecular weight= 27,523).

The start of the endonuclease gene was confirmed by N-terminal amino acid analysis of the purified endonuclease protein. The first 27 amino acids of the enzyme were found to correspond exactly with codons 2 thru 28 of the DNA sequence. The initial methionine residue predicted to be present from the DNA sequence was not found in the purified protein, suggesting that it is removed by processing. However, only 15% of the purified protein was able to be sequenced, and it is conceivable that in the remaining 85%, the N-terminal amino acid is blocked from Edman degradation by acetylation or formylation. The source of the enzyme used for sequencing was the *E. coli* clone, pSW149RM-3A. It is not known whether similar processing also occurs in *T. aquaticus*.

The start of the methylase gene has not been corroborated by protein sequencing. If it begins at the ATG codon at coordinate 673, it initiates

FIGURE 3: Nucleotide sequence of the *Taq* I restriction and modification genes and deduced amino acid sequence of the proteins. The positions of the *Taq* I recognition sites in the sequence are shown underlined. The N-terminal amino-acids of the endonuclease that were confirmed by protein sequencing are shown in italics.

an open reading frame of 363 codons which is predicted to specify a protein of molecular weight of 40,576. This agrees with the molecular weight of 40,000 as determined from its partial purification (see Materials and Methods). It is also possible that the modification gene starts 90 bases further downstream at the next ATG at coordinate 763.

The restriction gene is preceded by an extensive Shine-Dalgarno sequence (39), TAAGGAGG. There is no obvious Shine-Dalgarno sequence preceding the start of the modification gene if it starts at coordinate 673, although there is one in front of the ATG codon at position 763. There do not appear to be any recognizable *E. coli*-like promotor sequences within several hundred bases upstream of either the restriction gene or the modification gene. This would be consistent with the orientation dependent expression of the methylase gene. Both genes terminate with the stop codon TGA.

An endogenous promotor for restriction gene expression appears to be located within the methylase gene. There is a single *EcoR* I site in the 3' region of the methylase gene and a single *EcoR* I site in pBR322. Deletion of the DNA between the *Eco*R I sites of pSW149RM-4A resulted in derivatives which had lost most of the modification gene but which had retained the entire restriction gene. These derivatives were expected to behave like the reverse-orientation clones and to synthesize endonuclease but not methylase. In fact, they were found to synthesize neither enzyme at a detectable level, suggesting that part or all of the endonuclease promotor lies on the 5' side of the *Eco*R I site. Several candidate *E. coli*-like promotor sequences can be identified in this region but none are compelling. It seems reasonable to suppose that the methylase and endonuclease genes are preceeded by *T. aquaticus* promotors but that these are not recognized by *E. coli*, and that the sequence that does promote the transcription of the endonuclease in *E. coli* is fortuitous and possibly without function in *T. aquaticus*.

Strain-dependent viability

During the subcloning and manipulation of the *Taq* I restriction and modification clones, it became apparent that the *E.coli* strain used for the transformation seriously affected the outcome of the experiment. Strain RR1, the preferred host, was found to transform efficiently regardless of whether the plasmids synthesize methylase or endonuclease. Strain MM294, in contrast, was found to be highly sensitive to the methylase, and somewhat sensitive to the endonuclease. Sensitivity was observed in two ways: by a reduction in the transformation efficiency and by the distressed appearance of the transformants. The plasmids from the parental *BamH* I clones (M^+R^+) were found to transform MM294 at 10% of the expected frequency and the transformants appeared small, flat and translucent. The plasmids from the parental *Pst* I clones (M^+R^-) transformed MM294 at a higher frequency, but the colonies continued to appear distressed. The reverse-orientation *BamH* I clones (M^-R^+) transformed at normal frequency and the transformants appeared

normal in size, although slightly translucent. When both genes were eliminated (reverse-orientation *Pst* I clones, M^-R^-) transformants arose at a normal frequency and appeared normal (Table 1). MM294 appears to be sensitive to N^6-adenine methylation, at the sequences modified by the *Taq* I methylase. Several strains of *E. coli* are known to be sensitive to methylation and hydroxymethylation of cytosine (13,34,40,41) and at least two of the genes involved, *Mcr*A and *Mcr*B, have been identified. An analogous gene, *Mrr*, has been identified which sensitizes *E. coli* to adenine methylation (42) and it is probable that it is this function which is responsible for the sensitivity of MM294 to *Taq* I modification.

Another strain of *E. coli*, K802, displays greater sensitivity to both the *Taq* I methylase and endonuclease than does MM294. Plasmids that express both functions are unable to transform K802 at a detectable frequency. Transformation with plasmids that express either only the methylase or only the endonuclease occur at a reduced frequency and the colonies appear tiny and distressed. Plasmids that express neither gene transform normally (Table 1). The response of K802, therefore, is both more severe and more complex than MM294. Its behavior cautions that not all strains of *E. coli* are equally suitable for the cloning of restriction and modification genes.

G + C composition and codon utilization

T. aquaticus is a thermophile and its DNA might be expected to be G+C-rich. For the modification and restriction genes, the disproportion is slight: 59% G+C and 52% G+C respectively. For the sequence external to these genes, it is more substantial, 66% G+C. Codon utilization does not differ substantially from that used in *E. coli* (11). The third position

Table 1

Taq I clone	phenotype	*E. coli* strain		
		RR1	MM294	K802
BamH I parent (pSW149RM-3A)	(M^+R^+)	100% normal	10% small, flat	<0.1% tiny
BamH I Reverse (pSW149RM-3B)	(M^-R^+)	100% translucent	100% translucent	10% tiny
Pst I parent (pSW149M-4A)	(M^+R^-)	100% normal	50% small	25% tiny, flat
Pst I Reverse (pSW149M-4B)	(M^-R^-)	100% normal	100% normal	100% normal

TABLE 1: Relative transformation efficiencies of plasmids carrying the *Taq* I restriction and modification genes and the appearances of the transformants.

nucleotide usage appears to reflect the G+C content of *T. aquaticus* DNA as a whole, in that G and C are preferred.

Chromosomal location of the *Taq* I restriction-modification genes

Although a number of Type II restriction-modification systems are located on plasmids (7,8,10,11,13,43,44,45), the *Taq* I restriction-modification system is probably chromosomal. While *T. aquaticus* cells contain several plasmids which could be isolated by density gradient ultracentrifugation, blots of the nick-translated 5.5 kb *Bam*H I fragment were found to hybridize to high molecular weight *T. aquaticus* DNA and not to the plasmids. It remains possible, however, that the *Taq* I system resides on a large plasmid which could not be isolated during plasmid preparations and which co-purifies, and co-migrates, with chromosomal DNA.

DISCUSSION

The cloning of the *Taq* I restriction-modification system was accomplished by selecting *in vitro* for self-modified plasmids. This procedure was first suggested by Mann et al (6), and has been used on numerous occasions to clone both individual modification genes and complete restriction-modification systems from *Bacillus* and *Bacillus* phages (12,46,53,54,55,56,57,58,59), *Desulphovibrio* (14), *E. coli* and phage T4 (60,61,62), *Haemophilus* (19), *Moraxella* (63) and *Streptococcus* (15,64). An alternative isolation procedure, using phage to select *in vivo* for restricting clones, has also been used, but experience has shown it to be less reliable. The reason for the difference lies in the observation that many cloned systems modify well but restrict poorly. *Taq* I is an example of such a system: the parental *Bam*H I clones synthesize substantial levels of methylase and endonuclease, display full modification, yet remain completely sensitive to phage infection. The natural *E. coli* restriction-modification systems restrict phage severely (65,66,67,43,44), as do some of the foreign systems cloned into *E. coli*: for example *Hha* II (6), Dde I (14), *Pst* I (9), *Pae*R7 I (10) and *Pvu* II (13). It is not clear why the *Taq* I system fails to restrict; it could be due to inadequate endonuclease activity, to inappropriate cellular location, to the absence of required *T. aquaticus* factors like exonucleases, or to the possibility that its natural function in *T.aquaticus* is not bacteriophage-restriction at all.

The *Taq* I restriction-modification system exhibits a further unusual property in *E. coli*: the restriction gene is not lethal in the absence of modification. Several other cloned systems also behave in this way: for example, *Pae*R7 I (10), *Hae* II, *Hgi*A I, *Hinf* I, *Pst* I, and *Xba* I (Wilson *et al*, in preparation). Unmodified cells which contain these endonuclease genes differ strikingly from modified or normal cells: the colonies have a translucent, flat appearance compared to the normal, opaque, dome-shaped appearance. The cells experience trauma, as is evident by, 1) the high

frequency with which they are observed to lose the plasmid or to mutate to R^- and, 2) the low level of supercoiled plasmid DNA, and the high level of non-viscous chromosomal DNA fragments, present in cleared lysates during plasmid preparations. The extent to which restriction-modification clones are affected by the loss of the modification gene varies. The *Taq* I R^+ M^- clones appear to be the least affected, perhaps because the parental R^+ M^+ clones do not restrict exogenous phage DNA, and because the endonuclease normally functions at 65°C rather than at 37°C. The observation that the modification-deficient endonuclease-proficient clones are viable suggests that *E. coli* possesses an efficient mechanism for repairing endonucleolytic cleavage.

Comparing the sequence of the *Taq* I methylase with that of other published m^6A methylase sequences (7,11,17,46,47,48,49,50,51,52, Murray *et al*, in press, Chandrasegaran and Smith, personal communication) reveals that the closest relative of *Taq* I is *Pae*R7 I (50). Since the *Pae*R7 I recognition sequence (CTCGAG) is a subset of the *Taq* I recognition sequence, the homology between the two methylases might reflect common elements determining sequence specificity. A similar comparison of the sequence of the *Taq* I methylase sequence with the published sequences of m^5C methylases revealed no substantial homologies, reinforcing the observation that m^6A and m^5C methylases have distinctly different structures. The sequence of the *Taq* I endonuclease was compared with published sequences of other Type II endonucleases and no substantial homologies were detected.

An interesting feature of the *Taq* I restriction-modification system concerns the uneven distribution of *Taq* I sites between the two genes. There are seven Taq I sites in the restriction gene and none in the modification gene. Given the base compositions and lengths of the genes, and assuming a random base distribution, the expected numbers of sites in the restriction and modification genes are 3 and 4 respectively. The probability of 7, or more, sites occurring in the restriction gene is 2%, and the probability of no sites occurring in the modification gene is also 2%. It is unlikely that these situations would occur individually and less likely that they would occur together (Chi-squared = 10.8, $P < 1\%$). It is possible that the distribution of sites has a regulatory function. Thus, in an under-modified cell interaction of the endonuclease with sensitive *Taq* I sites in the restriction gene might interrupt transcription of the gene, and prevent further endonuclease synthesis until the cell had become fully modified. Such a mechanism could facilitate the acquisition of the system by a new host and could enable a cell to cope with subsequent variations in the degree of modification.

The key feature of a regulatory mechanism of this type is the presence of self-specific sites in the restriction gene and their absence in the

modification gene. Analysis of published sequences for modification and restriction genes shows that this feature is not widespread; in fact, it currently occurs in no other systems but *Taq* I. It is clear that auto-restriction cannot be a common method for regulating endonuclease expression. It remains to be proven whether it works in the *Taq* I system and, if it does, whether it provides a safety measure that reduces the likelihood of the system killing its host. Whether such a system could function in a more general way, so as to regulate the expression of other genes that possess many *Taq* I sites, also remains to be seen.

ACKNOWLEDGEMENTS

We wish to thank J. Beckwith, J. Brooks, D. Comb, C. Guan, N. Kleckner, M. Nelson, E. Raleigh, R. Roberts, E. Rosenvold, and I. Schildkraut for constructive criticism and comments throughout this research. We wish to also thank F. Barany, N. Murray and H. Smith for many discussions and for sharing their unpublished results with us.

*To whom correspondence should be addressed

REFERENCES

1. Roberts, R. J. (1987) Nucleic Acids Res.15, r189-r217.
2. Xia, Y., Burbank, D.E. and van Etten, J.L. (1986) Nucleic Acids Res. 14, 6017-6030.
3. Sato, S., Hutchison, C. A. III and Harris, J. I (1977) Proc. Natl. Acad. Sci. USA 74, 542-546.
4. Barker, D., Hoff, M., Oliphant, A. and White, R. (1984) Nucleic Acids Res. 12, 5567-5581.
5. Sato, S., Nakazawa, K. and Shinomiya, T. (1980) J.Biochem. 88, 737-747.
6. Mann, M. B., Rao, R. N. and Smith, H. O. Gene (1978) 3, 97-112.
7. Newman, A. K., Rubin, R. A., Kim, S-H. and Modrich, P. (1981) J. Biol. Chem. 256, 2131-2139.
8. Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E. and Rosenberg, J.M. (1981) J. Biol. Chem. 256, 2143-2153.
9. Walder, R. Y., Hartley, J. L., Donelson, J. E. and Walder, J. A.(1981) Proc. Natl. Acad. Sci. USA 78, 1503-1507.
10. Gingeras, T. R. and Brooks, J. E. (1983) Proc. Natl. Acad. Sci. USA 80, 402-406.
11. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. Nucleic Acids Res. (1984) 12, 3659-3676.
12. Kiss, A., Posfai, G., Keller, C. C., Venetianer, P. and Roberts, R. J. (1985) Nucleic Acids Res.13, 6403-6421.
13. Blumenthal, R. M., Gregory, S. A. and Cooperider, J. S.(1985) J. Bacteriol. 164, 501-509.
14. Howard, K. A., Card, C., Benner, J. S., Callahan, H. L., Maunus, R., Silber, K., Wilson, G. and Brooks, J. E. (1986) Nucleic Acids Res. 14, 7939-7951.
15. Lacks, S. A, Mannarelli, B. M., Springhorn, S. S. and Greenberg, B. (1986) Cell 46, 993-1000.
16. Kosykh, V. G., Buryanov, Y. I. and Bayev, A. A. (1980) Molec. Gen. Genet. 178, 717-718.

17. Sznyter, L. A., Slatko, B., Moran, L., Howard, K., and Brooks, J. E. (1987) Nucleic Acids Res.(in press)
18. Cote, M. A., Daggett, P. M., Gantt, M. J., Hay, R., Jong, S. and Pienta, P. (1984) American Type Culture Collection Media Handbook, Rockville, Md.
19. Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G. G. and Wells, R. D. (1987) J. Biol. Chem. 262, 4770-4777.
20. Clewell, D. B. and Helinski, D. R. (1969) Proc. Natl. Acad.Sci. USA 62, 1159-1166.
21. Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
22. Mandel, M. and Higa, A. (1970) J. Mol. Biol. 53, 159-162.
23. Lederberg, E. M. and Cohen, S. N. (1974) J.Bacteriol. 119, 1072-1074.
24. Miller, J. (1972) Experiments In Molecular Genetics. Cold Spring Harbor Laboratory. Cold Spring Harbor, N. Y.
25. Sanger, F.,Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
26. Williams, S. A., Slatko, B. E., Moran, L. S., Simone, S.A. (1986) BioTechniques 4, 138-147.
27. Chen, E.J. and Seeburg,P. H. (1985) DNA 4, 165-170.
28. Haltiner, M., Kempe, T. and Tjian, R. (1985) Nuc. Acids Res. 13, 1015-1019.
29. Hattori, M.and Sakaki, Y. (1986) Anaylyt. Biochem. 152, 232-238.
30. Biggin, M. D., Gibson, T. J. and Hong, G. F (1983) Proc. Natl.Acad Sci. USA 80, 3963-3965.
31. Ansorge, W. and Labeit, S. (1984) J. Biochem. Biophys. Methods 10, 237-243.
32. Olsson, A., Moks, T., Muhlen, M. and Gaal, A. B. (1984) J. Biochem. Biophys. Methods 10, 83-90.
33. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
34. Raleigh, E. A and Wilson, G. (1986). Proc. Natl. Acad. Sci. USA 83, 9070-9074
35. Keller, C., Corcoran, N. and Roberts, R. E. (1984) Nucleic Acids Res.12, 379-386.
36. Devereux, J., Haeberli, P. and Smithes, O. (1984) Nucl Acids Res. 12, 387-395.
37. Strickler, J. E., Hunkapiller, M. W. and Wilson, K. J. (1984) Analytical Biochemistry 140, 553-566.
38. Hunkapiller, M. W. and Hood, L.E. (1983) Methods in Enzymology 91, 486-493
39. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
40. Revel, H. R. and Luria, S. E. (1970) Ann. Rev. Genetics 4, 177-192.
41. Ross, T. K. and Braymer, H. D. (1987) J.Bacteriol. 169, 1757-1759.
42. Heitman, J. and Model, P. (1987) J. Bacteriol. 169, (in press).
43. Mise, K. and Nakajima, K. (1985) Gene 33, 357-361.
44. Yoshida, Y. and Mise, K. (1986) J.Bacteriol. 165, 357-362.
45. Som, S., Bhagwat, A, S. and Friedman, S. (1987) Nucleic Acids Res. 15: 313-332.
46. Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) Nucleic Acids Res. 12, 9039-9049.
47. Walder, R. Y., Walder, J. A. and Donelson, J. E. (1984) J. Biol. Chem. 259, 8015-8026.
48. Mannarelli, B. M., Balganesh, T. S., Greenberg, B., Springhorn, S. S. and Lacks, S. A. (1985) Proc. Natl. Acad. Sci. USA 82, 4468-4472.
49. Macdonald, P. M. and Mosig, G. (1984) EMBO J. 3, 2863-2871.
50. Theriault, G., Roy, P. H., Howard, K .A., Benner, J. S., Brooks, J. E, Waters, A. F.and Gingeras, T. R. (1985) Nucleic Acids Res. 13, 8441-8461.
51. Schoner, B., Kelley, S. and Smith, H. O. (1983) Gene 24, 227-236.

52. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., McDonald, P. M. and Mosig, G. (1985) J. Bacteriol. 164, 932-937.
53. Szomolanyi, E., Kiss, A. and Venetianer, P. (1980) Gene 10, 219-225.
54. Janulaitis, A., Povilionis, P. and Sasnauskas, K. (1982) Gene 20 (1982) 197-204.
55. Kiss, A. and Baldauf, F. (1983) Gene 21, 111-119.
56. Behrens, B., Pawlek, B., Morelli, G. and Trautner, T. A. (1983) Mol. Gen. Genet. 189, 10-16.
57. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T. S. and Trautner,T. A. (1987) EMBO J. 6, 1137-1142.
58. Gunthert, U., Reiners, L. and Lauster, R. (1986) Gene 41, 261-270.
59. Noyer-Weidner, M., Jentsch, S., Kupsch, J., Bergbauer, M. and Trautner, T. A. (1985) Gene 35, 143-150.
60. Herman, G. E. and Modrich, P. (1982) J.Biol.Chem. 257, 2605-2612.
61. Brooks J. E., Blumenthal, R. M. and Gingeras, T. R.(1983) Nucleic Acids Res. 11, 837-851.
62. Schlagman, S. L. and Hattman, S. (1983) Gene 22, 139-156.
63. Walder, R. Y., Langtimm, C. J., Chatterjee, R. and Walder, J. A. (1983) J. Biol. Chem. 258, 1235-1241.
64. Lacks, S. A. and Springhorn, S. S. (1984) J.Bacteriol. 157, 934-936.
65. Arber, W. (1971) The Bacteriophage Lambda (A.D.Hershey, ed.) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp 83-96.
66. Smith, H. O. and Kelly, S. V. (1984) In, DNA Methylation, Biochemistry and Biological Significance (A. Razin, H. Cedar, A. D. Riggs, eds.). Springer-Verlag, N.Y. pp 39-71.
67. Yuan, R. and Hamilton, D. L. (1984) In, DNA Methylation, Biochemistry and Biological Significance (A. Razin, H. Cedar, A. D. Riggs, eds.) Springer-Verlag, N. Y. pp 11-37.