Volume 13 Number 23 1985 Nucleic Acids Research

Nucleotide sequence of the *Pae*R7 restriction/modification system and partial characterization of its protein products

G.Theriault[1], P.H.Roy[1], K.A.Howard[2], J.S.Benner[2], J.E.Brooks[2], A.F.Waters and T.R.Gingeras

La Jolla Biological Laboratories, P.O. Box 85350, San Diego, CA 92138-9216, USA, [1]Department of Biochemistry, Faculty of Sciences, Laval University, Ste.-Foy, Quebec G1K 7P4, Canada, and [2]New England BioLabs, 32 Tozer Road, Beverly, MA 01915, USA

Received 12 August 1985; Revised and Accepted 8 November 1985

ABSTRACT

*Bal*31 deletion experiments on clones of the *Pae*R7 restriction-modification system from *Pseudomonas aeruginosa* demonstrate that it is arranged as an operon, with the methylase gene preceding the endonuclease gene. The DNA sequence of this operon agrees with *in vitro* transcription-translation assays which predict proteins of 532 amino acids, Mr = 59,260 daltons, and 246 amino acids, Mr = 27,280 daltons, coincident with the methylase and endonuclease genes, respectively. These predicted values coincide with the measured molecular weights of the purified, denatured *Pae*R7 endonuclease and methylase proteins. The first twenty amino acids from the amino-terminus of the purified endonuclease exactly match those predicted from the DNA sequence. Finally, potential regulatory mechanisms for the expression of phage restriction are described based on the properties of several *Pae*R7 subclones.

INTRODUCTION

Restriction-modification systems provide a means for host bacterial cells to recognize and destroy foreign DNA. Of the three types of prokaryotic restriction-modification systems (1-3), the genes of the type II systems have been most frequently studied because of their simplicity and their production of endonucleases commonly used in genetic engineering. Of the type II systems which have been cloned and characterized, the *Hha*II (4,5), *Pst*I (6,7) and *Msp*I (R. Roberts, personal communication) systems are believed to be located on the bacterial chromosomes, while the *Eco*RI (8,9), *Eco*RII (10), *Eco*RV (11), *Pvu*II (12) and *Pae*R7 (13,14) systems are encoded on naturally occurring bacterial plasmids.

The *Pae*R7 system was originally identified as part of a 42 kilobase plasmid, pMG7, found in *Pseudomonas aeruginosa* (15). After being identified as a type II system by Hinkle and Miller

© IRL Press Limited, Oxford, England.

(16), the *Pae*R7 system was transferred onto *E. coli* (13) and subcloned and characterized on a 3.8 kilobase DNA fragment (pPAORM3.8) from pMG7 (14). The *Pae*R7 system is unique among characterized restriction-modification systems in its ability to permit the separation of the endonuclease and methylase genes into viable clones (14). The expected lethality of an expressing endonuclease gene in an unmodified host was not observed with clones carrying the *Pae*R7 endonuclease. In addition, *E. coli* cells expressing only the *Pae*R7 endonuclease no longer appeared capable of restricting bacteriophage. The questions of how unmodified host cells expressing the endonuclease survive and of why bacterial cells containing a restriction endonuclease were no longer capable of restricting infecting bacteriophage called for additional investigation.

In an attempt to answer these questions, we have begun by determining the structure and organization of the *Pae*R7 genes as well as partially characterizing their gene products. In a companion paper (17), we describe how the structure and expression of these genes affects the viability of the host cell and its ability to restrict infecting bacteriophage.

MATERIALS AND METHODS

Phage and DNA Preparations:

Plasmid DNA used in the experiments was isolated by the cleared lysate method (18) followed by banding in CsCl/ethidium bromide gradients. Small scale preparations of plasmid DNA from clones were obtained by a modified version of Birnboim and Doly (19). High titer stocks of ø80 phage for *in vivo* restriction analysis were prepared by plate lysates (20).

E. coli strains MM294 and RR1 [20] were used for cloning and enzyme purification. JM103 was used for M13 cloning using M13 vectors mp7, mp8, mp9, mp18 and mp19 (21,22).

Clone Constructions:

Deletions in clones carrying the *Pae*R7 methylase and endonuclease genes were constructed by use of *Bal*31 nuclease (Bethesda Research Laboratories) as described previously (14). Restriction endonuclease mapping of all constructions was done using enzymes from New England BioLabs following conditions

recommended by the manufacturer. E. coli cells were transformed by the $CaCl_2$ heat shock method of Cohen et al. (23).

A plasmid which expresses elevated levels of the endonuclease and methylase proteins was constructed in the following manner (Figure 1). A SalI linker was inserted into pPAORM3.8 linearized at the NruI site. Five μg of this new construct (designated pPAORM3.8[S]) and 5 μg of the lambda P_L promoter regulated expression vector pGW7 (from G. Wilson) were each digested in a reaction volume of 100 μl using an excess of SalI; the reactions were stopped by incubation at 65°C for 10 min. pGW7 was then digested by an excess of BamHI. pPAORM3.8[s] was digested with an excess of PstI, PvuII, BamHI and SalI for 2 hours at 37°C. After phenol extraction and ethanol precipitation, 1 μg of each digest was combined and ligated in 100 μl. The ligation mix was used to transform RR1 cells; transformants were selected for Ap(r) and individual colonies tested for PaeR7 endonuclease activity.

Analysis of restriction-modification activities:

In vivo restriction-modification tests have previously been described (14). Similarly in vitro assays of cell extracts were performed as described except Ø80 DNA was substituted for Adenovirus-2 DNA.

DNA Sequencing:

A portion of the DNA sequencing was performed according to the chemical cleavage method of Maxam and Gilbert (24). Plasmids pPAORM3.2, pPAOM2.7, pPAO2.2 and pPAO1.6 were end-labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase (New England BioLabs) at the SalI linker sites to determine the junction between the vector and Bal31-treated insert. In the plasmids pPAOΔ594, pPAOΔ614, pPAOΔ628 and pPAORM3.8, a BstNI site approximately 100 base pairs away from the junction sequences for each of the deletion clones (Figure 3) was cleaved, end-labeled with ^{32}P and used to determine the DNA sequences.

The dideoxynucleotide chain-termination procedure (25) was also used to determine the sequence of the PaeR7 operon. Templates were prepared from M13 clones which contained either various restriction fragments from the insert in pPAORM3.8 or the entire 3800 base pair PaeR7 operon inserted into the vectors

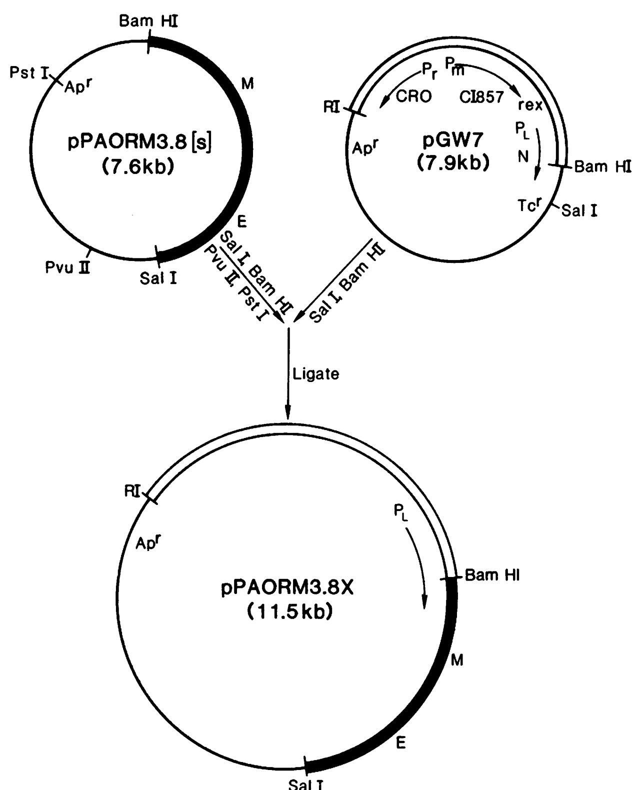


Figure 1. CONSTRUCTION OF pPAORM3.8-X, A PLASMID WHICH OVER-PRODUCES ENDONUCLEASE AND METHYLASE.

mp7, mp8, mp9, mp18 or mp19 (21,22). In addition to the 17 nucleotide universal primer (New England BioLabs) used in the M13 sequencing system, a collection of 10 oligonucleotides complementary to various regions within the 3800 base pair PaeR7 insert was synthesized (using phosphoramidite chemistry in an Applied Biosystems Synthesizer) and used as internal primers to elucidate about 1540 nucleotides of the PaeR7 system. DNA sequences were resolved using 6, 8 and 12% thin polyacrylamide gels run in a gradient buffer system (26). ^{35}S and ^{32}P-labeled nucleotides were used to visualize the DNA sequences.

Protein sequencing:

A Waters Associates Liquid Chromatograph was used for C4 and Cyano reverse phase chromatography. Protein samples were subjected to a final chromatography on a Vydac C4 214TP54 (5 µm,

4.6 x 300 mm) 300 Angstrom pore reverse phase column developed with a linear gradient of 5% to 50% acetonitrile in 0.1% trifluoroacetic acid over 25 min. at a flow rate of 1 ml/min 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. The phenylthiohydantoins were unambigiously identified by high-performance liquid chromatography on an IBM Cyano (5 μm, 4.5 x 250 mm) column with slight gradient modifications.

In vitro Transcription-Translation:

Experiments involving *in vitro* transcription-translation of plasmids encoding all or part of the *Pae*R7 restriction-modification system were performed by using kits supplied by Worthington, Inc. The protocol used was the one recommended by the manufacturer. Reactions included using 3 μg of supercoiled DNA for plasmids pPAORM3.8, pPAOR1.9, pPAOM2.7 and pBR322. In addition, 3 μg of each of these plasmid DNAs were cleaved with *Hinc*II to block the expression of the beta-lactamase protein so as to visualize the production of the 27,000 dalton *Pae*R7 endonuclease. Proteins were labeled with ^{35}S-methionine as a result of the translation process and resolved using an 11% polyacrylamide and SDS gel with a 5% stacking gel (27).

Isolation and Characterization of Methylase and Endonuclease:

The plasmid pX164 DNA linearized with *Pvu*II (pX164/*Pvu*II) and was used as a substrate for the detection of *Pae*R7 endonuclease activity. pX164 is a pBR322 derivative containing a *Xho*I linker (CCTCGAGG) ligated at position 164 in the pBR322 sequence (14). Endonuclease and methylase enzymes were diluted into *Pae*R7 storage buffer, as recommended by New England BioLabs.

During the purification, column fractions were assayed in a 50 μl reaction containing *Pae*R7 assay buffer and 1 μg pX164/*Pvu*II. 1 μl column samples were added to the reaction. To quantify endonuclease activity, extracts were serially diluted in 10-fold increments, and reactions were incubated for 20 min. at 37°C. One unit of *Pae*R7 endonuclease is defined as the amount of enzyme required to cut pX164/*Pvu*II DNA to completion.

Column fractions were assayed for methylase activity, using ligated *Xho*I linkers as substrate, in a 25 μl reaction volume

containing 50 mM Tris, pH 7.5; 10 mM EDTA, 5 mM beta-mercapto-ethanol, 100 µg/ml BSA, 80 µM [^{3}H]S-Adenosyl-L-methionine [New England Nuclear, 12.8 Ci/mmol] and 0.5 µg ligated XhoI linkers. Reactions were incubated for 1 hour, 37°C, then stopped by the addition of an equal volume of 10% TCA. The precipitated reactions were spotted onto Whatman 5 MM filter paper. The filter paper was submerged and shaken at 4°C in 10% TCA, washed one time with more 10% TCA, rinsed in isopropanol and dried. The filter paper was cut into 1-inch squares and counted, using 2 ml of opti-fluor [Packard], in a Packard Tricarb scintillation counter.

Quantitation of methylase units was done by a protection assay in which 1 µg of pX164/PvuII plasmid served as a substrate in a 50 µl reaction volume. Reaction components were the same as above, except 1 mM unlabeled S-Adenosyl-L-methionine [Sigma] was used as the methyl donor. Reactions were incubated for 15 minutes at 37°C, phenol extracted, ethanol precipitated, resuspended in 1x XhoI buffer (New England BioLabs) and challenged with 10 units of XhoI (or PaeR7) enzyme for 30 minutes at 37°C. The reaction products were resolved by agarose gel electrophoresis. One unit of PaeR7 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 pX164/PvuII DNA against PaeR7 (or XhoI) cleavage.

Quantitative protein determinations were done using the protein dye binding assay of Bradford (28) with bovine serum albumin as a standard.

Protein Purifications and Characterization:

Cell growth and soluble extract preparation: *E. coli* RR1 cells containing the pPAORM3.8-X overproducing clone were grown to A_{550}=0.9 at 30°C in L-broth (24 liters). The cells were induced by shifting the temperature to 43°C, and 82 grams were harvested by centrifugation three hours later.

Forty-one grams of cells were resuspended in 100 ml of Buffer S (10 mM Tris, pH 7.6; 10 mM beta-mercaptoethanol). Cells were partially lysed by freezing at -70°C and thawing in a solution containing 500 µg/ml lysozyme and 10 mM EDTA, pH 7.9. Thawed cells were sonicated at full intensity for two 1-minute bursts. After bringing the suspension to 0.5 M NaCl, sonicated

cells were centrifuged at 10,000 x g in a Type JA-14 rotor (Beckman) for 10 minutes at $4^{o}C$.

The resulting supernatant (131 ml) was made 7.5% with PEG 6000. After precipitating for one hour at $4^{o}C$, this solution was centrifuged at 10,000 x g for 30 minutes. The 17 gram pellet was resuspended in approximately 500 ml of Buffer S using a Waring blender. The small portion of the pellet that did not redissolve (<1%) was eliminated by centrifugation. All subsequent purification steps took place at $0-4^{o}C$, unless otherwise stated.

Heparin-Sepharose Chromatography: The 500 ml resuspended PEG pellet was loaded on a Heparin-Sepharose (Pharmacia) column (2.5 x 14 cm) equilibrated and washed with Buffer SN (Buffer S and 50 mM NaCl). The enzyme was eluted with a linear gradient of 700 ml of Buffer S containing 50 mM to 1.0 M NaCl, and 6.5 ml fractions were collected. The PaeR7 endonuclease eluted at approximately 0.2 M NaCl. Peak fractions were pooled, and the 40 ml peak was dialyzed against 2 liters of Buffer Q (20 mM Tris, pH 7.5; 10 mM beta-mercaptoethanol). The PaeR7 methylase eluted as a single peak at approximately 0.33 M NaCl. Peak fractions were pooled, and the 33 ml peak was dialyzed against KPO_4 Buffer (20 mM KPO_4, pH 6.9; 10 mM beta-mercaptoethanol).

Synchromprep Q-300 Chromatography (for endonuclease): Because of high protein concentrations, the PaeR7 endonuclease pool was divided into two batches. Each was loaded on a 1 x 8 cm column packed with Synchromprep Q-300, 30 μm, (Synchrom, Inc., Linden, Indiana) equilibrated and washed with 50 mM KCl in Buffer Q. A gradient from 50 mM to 0.9 M KCl was used; the endonuclease eluted between 0.6 M and 0.7 M KCl. Peak fractions were pooled into a 300 ml total and dialyzed against KPO_4 buffer.

PolyCAT A Chromatography (for endonuclease): A 9.4 x 250 mm polyCAT A column (Custom LC, Inc.) was charged with 1.0 M KCl in KPO_4 Buffer and then equilibrated with 50 mM KCl. Two 150 ml portions of the pooled fractions from the Synchroprep column were loaded and developed independently. The column was washed with 50 mM KCl/KPO_4 Buffer and the enzyme eluted with a 75 ml linear gradient from 50 mM to 1.0 M KCl. The PaeR7 endonuclease eluted at approximately 0.33 M KCl and was stored in

TABLE 1
PURIFICATION OF PaeR7 ENDONUCLEASE

Purification Step	Total Protein (mg)	Total Activity (Units)	Yield (%)	Specific Activity (U/mg)
Sonicated Crude Extract	1800	1×10^9	100	5.5×10^5
PEG Pellet	1000	5×10^8	50	5×10^5
Heparin Sepharose	28.6	1.3×10^8	13	4.5×10^6
Synchroprep	24[a]	1.2×10^8[a]	12	5×10^6
PolycatA	12	1×10^8	10	6.2×10^6

[a]Values extrapolated from OD profile of enzyme elution.

37% glycerol at either -20°C or -70°C.

PolyCAT A Chromotagraphy (for methylase): 32 ml of PaeR7 methylase from the Heparin-Sepharose column was loaded onto a PolyCAT A column prepared as described above. The enzyme was eluted with a linear gradient of 48 ml from 50 mM to 0.6 M KCl. The methylase eluted at 0.4 M KCl in 4 ml. This pool was diluted to 28 ml in Buffer Q, bringing the solution to approximately 50 mM KCl.

Polyanion SI (for methylase): The 28 ml pool of methylase was loaded on a 0.5 x 5 cm Polyanion-SI-17 column (Pharmacia), and the column was washed with 50 mM KCl/Buffer Q. The enzyme was eluted by running a 46 ml linear gradient from 50 mM to 0.9 M KCl, with the PaeR7 methylase eluting at 0.15 M KCl. A 1 ml fraction contained the majority of the methylase. The methylase was stored in 37% glycerol at -20°C. The enzymatic activity of the methylase and endonuclease at each step in the purification was determined, and the latter is detailed in Table 1.

Molecular Weight Determination by Gel Filtration: Molecular weights of the native form of the endonuclease and methylase proteins were determined by gel filtration using a 2.5 x 78 cm Sephadex G-100 column (Pharmacia). The calibration standards (Bovine Serum Albumin, Ovalbumin, and Ribonuclease A) were prepared and applied to the Sephadex G100 column according to the manufacturer.

The endonuclease and methylase samples used were taken from Heparin Sepharose peaks. 1.5 ml (1.5×10^5 units) of endonu-

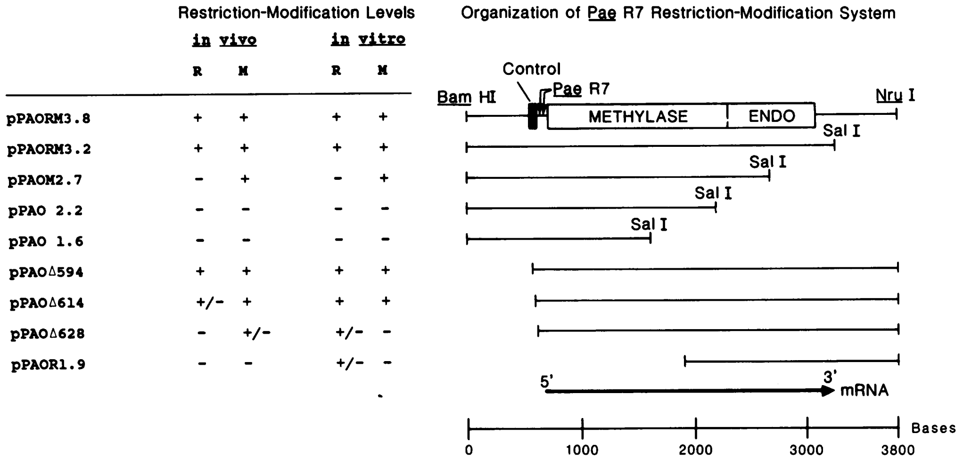


Figure 2. SUMMARY OF THE BAL31 DELETION MUTANTS DERIVED FROM pPAORM3.8. Each mutant was tested by in vitro and in vivo assays for restriction modification phenotype. (+)= an activity level equivalent to pPAORM3.8. (+/-)= a 10x or greater decrease in activity relative to pPAORM3.8. (-)= no detectable activity. Bal31 deletion mutants were named as previously described (14), except pPAOΔ594, pPAOΔ614 and pPAOΔ628, where the number following the "Δ" indicates the approximate distance of the deletion end point from the BamHI site.

clease was applied to the column. 4.7 ml fractions were collected and assayed for enzymatic activity as previously described, using 1 µl from each fraction. 1.5 ml (1.0×10^4 units) of methylase was applied to the column. 3.7 ml fractions were collected and tested with the protection assay, using 10 µl from each fraction. The K_{av} was also calculated for the standards, endonuclease and methylase, and these were plotted to determine the relative molecular weights.

RESULTS

Organization and DNA Sequence of PaeR7 System:

In the initial report concerning the cloning of the PaeR7 system, the order of the genes encoded in the 3800 base pair fragment was determined to be the methylase gene nearest to the BamHI site, with the endonuclease closest to the NruI site (Figure 2). To determine the boundaries of each of these genes more accurately within the 3800 base pair insert in pPAORM3.8, a second series of deletions was made from either the BamHI or

CGCCGACGTGGCGAAACGCTATGGCGTTTACACGGCCACCATTTACAAGCACGTTGGTGCGGTGCGCCCAGAGCGCGGCGCATCGCCGATGATTTAGTGAGGTCGTCATTCATGGCATTT
10 20 30 40 50 60 70 80 90 100 110 120

Δ 594 Δ614 Δ628
GCGCCAAGTGTTGCTCATAAGCCTGTTGCCGCTGCTGTATGCCCTGTCATGGCCGCGACCGAGGCGTTGGCAACGGAAGGGGGGCTCGAGGCTCGAGGGGCTATTTTCACGCGCTCCGAG
130 140 150 160 170 180 190 200 210 220 230 240

MET VAL ASP PHE ILE LEU ASP LEU ALA GLY TYR THR GLU ASP GLN PRO LEU HIS GLU LYS ARG LEU LEU GLU PRO SER PHE GLY GLY GLY ASP PHE LEU LEU PRO ILE ILE GLN ARG LEU
GTGGTCGATTTCATCCTCGACCTGGCTGGCTACACCGAGGATCAACCGCTGCACGAAAAGCGGCTTTTGGAACCGTCGTTCGGTGGCGGGGACTTCTTGCTGCCGATCATTCAGCGGCTG
250 260 270 280 290 300 310 320 330 340 350 360

LEU SER ALA TRP ARG ALA ALA ARG PRO ASN GLY THR GLU VAL ASP ASP LEU GLY ASP ALA ILE ARG ALA VAL GLU LEU HIS HIS ASP THR PHE ARG SER THR TYR ALA ALA VAL VAL ALA
CTGAGTGCATGGCGAGCAGCAAGGCCCAATGGTACTGAGGTCGATGATTTGGGCGACGCCATCCGGGCGGTGGAGTTGCACCACGACACCTTCCGCAGCACCTACGCTGCCGTTGTCGCG
370 380 390 400 410 420 430 440 450 460 470 480

LEU LEU LYS ARG GLU GLY LEU SER ALA ASN ALA ALA THR ALA LEU ALA ASP ARG TRP LEU SER GLN GLY ASP PHE LEU LEU ALA PRO LEU GLU GLY GLN PHE ASP PHE VAL VAL GLY ASN
CTGCTCAAGCGTGAGGGGCTGTCGGCGAATGCTGCGACGGCCTTGGCTGACCGCTGGCTGTCGCAGGGTGATTTCCTGCTGGCTCCGCTGGAGGGCCAGTTCGATTTCGTGGTGGGCAAT
490 500 510 520 530 540 550 560 570 580 590 600

PRO PRO TYR VAL ARG PRO GLU LEU ILE PRO ALA PRO LEU LEU ALA GLU TYR ARG SER ARG TYR GLN THR MET TYR ASP ARG ALA ASP ILE TYR ILE PRO PHE ILE GLU ARG SER LEU THR
CCTCCCTACGTTCGACCTGAGCTGATTCCGGCCCCTTTGCTGGCCGAGTACCGCAGCCGCTATCAGACGATGTATGACCGGGCGGACATCTACATTCCCTTCATCGAGCGGTCGCTGACG
610 620 630 640 650 660 670 680 690 700 710 720

ALA LEU SER ALA GLY GLY ASN LEU GLY PHE ILE CYS ALA ASP ARG TRP MET LYS ASN ARG TYR GLY GLY PRO LEU ARG SER LEU VAL ALA GLU ARG PHE HIS LEU LYS VAL TYR VAL ASP
GCATTGTCTGCTGGCGGCAATCTGGGCTTTATCTGCGCGGATCGCTGGATGAAGAACCGCTACGGTGGGCCACTGCGTAGCCTTGTTGCCGAACGGTTCCACCTGAAAGTCTATGTCGAT
730 740 750 760 770 780 790 800 810 820 830 840

MET VAL ASP THR PRO ALA PHE HIS SER ASP VAL ILE ALA TYR PRO ALA ILE THR ILE ILE SER ARG GLU GLY GLY GLY ALA THR ARG ILE ALA HIS ARG PRO SER ILE ASP ARG ALA THR
ATGGTGGATACACCGGCCTTCCATTCCGATGTGATCGCCTATCCGGCCATCACCATCATCAGCCGTGAGGGGGGCGGGGCGACGCGCATCGCACACCGCCCGTCCATAGACCGGGCCACG
850 860 870 880 890 900 910 920 930 940 950 960

LEU THR THR LEU ALA GLY LEU LEU SER ALA PRO THR LEU PRO LYS ASP ALA GLY PRO VAL ARG GLU LEU ALA ARG VAL THR ASN GLY ALA GLU ARG GLY CYS TRP SER LEU LEU THR ARG
CTGACCACGCTGGCTGGTCTGCTGTCGGCACCGACACTTCCGAAGGATGCTGGCCCGGTGCGGGAACTCGCCCGTGTGACCAATGGTGCGGAGCGTGGCTGCTGGAGTCTTCTGACCAGA
970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080

TRP ARG LEU PHE ALA VAL TRP ARG ALA ARG SER HIS CYS SER LYS ARG LEU GLY ALA ARG PHE GLY ILE GLY VAL ALA THR GLY ALA ASP LYS ALA PHE ILE GLY ASP PHE GLU SER LEU
TGGCGCTTATTCGCCGTCTGGAGGGCGCGTTCCCACTGCTCGAAGAGGCTGGGTGCAAGGTTCGGCATTGGTGTTGCGACGGGCGCTGACAAGGCATTCATCGGCGACTTCGAGTCGCTG
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200

ASP VAL GLU PRO ASP ARG LYS LEU PRO LEU VAL THR THR LYS ASP ILE MET THR GLY GLU VAL GLN TRP ARG GLY GLN GLY VAL ILE ASN PRO PHE ALA GLU SER GLY GLY LEU VAL ASP
GATGTCGAGCCTGATCGGAAGCTGCCGCTGGTGACAACCAAAGACATCATGACCGGCGAAGTGCAGTGGCGTGGTCAGGGGGTCATCAACCCCTTCGCCGAGTCTGGAGGGCTGGTCGAT
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320

LEU GLY GLU TYR PRO ARG LEU ARG ARG TYR LEU GLU ALA ARG ARG ASP VAL ILE ALA GLY ARG HIS CYS ALA LYS LYS ALA PRO ALA ASN TRP TYR ARG THR ILE ASP ARG ILE THR PRO
CTTGGCGAGTATCCGCGTCTGCGCCGCTACCTTGAGGCTCGCCGGGATGTGATTGCCGGTCGGCATTGCGCGAAAAAGGCTCCTGCCAACTGGTATCGCACGATTGACCGTATTACCCCG
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440

ALA LEU ALA ALA ARG PRO LYS LEU LEU ILE PRO ASP ILE LYS GLY GLU SER HIS ILE VAL PHE GLU GLY GLY GLU LEU TYR PRO SER HIS ASN LEU TYR TYR VAL THR SER ASP ASP TRP
GCGCTGGCTGCGAGGCCCAAGCTGTTGATCCCTGACATCAAGGGTGAGTCGCACATTGTTTTCGAGGGCGGCGAGCTGTATCCAAGCCACAACCTCTACTACGTTACCTCGGACGATTGG
1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560

ASP LEU ARG ALA LEU GLN ALA VAL LEU LEU SER ALA VAL SER ARG LEU PHE VAL ALA THR TYR SER THR LYS MET ARG GLY GLY PHE LEU ARG PHE GLN ALA GLN TYR LEU ARG ARG ILE
GATTTGCGGGCCTTGCAAGCTGTGCTGCTGTCCGCTGTCTCGCGCCTGTTCGTGGCCACGTACTCAACGAAAATGCGAGGTGGCTTCTTGCGCTTCCAAGCGCAATACCTGCGCCGCATC
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680

ARG ILE PRO ARG TRP ALA ASP VAL PRO GLU PRO LEU ARG ARG GLU LEU ALA GLU ALA ALA ILE LYS ARG ASP VAL GLN ALA CYS ASN ARG ALA VAL PHE ARG LEU TYR GLY LEU SER HIS
CGTATCCCACGCTGGGCCGATGTGCCTGAACCGCTACGCCGCGAGCTGGCTGAGGCAGCCATCAAGCGGGACGTGCAAGCCTGCAACCGGGCTGTGTTCAGGCTCTACGGGCTGAGCCAC
1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

GLU GLU ARG SER ALA LEU GLY GLY ASN GLY GLU MET ALA LEU ASP LEU VAL ASP TYR GLU GLN LYS ALA ARG ASP ALA VAL LYS ALA PHE TRP GLY ASN ARG GLU ALA ALA ARG GLN
GAAGAACGATCTGCCCTTGGGGGCAATGGAGAATAAATGGCACTTGATCTTGTCGATTACGAACAGAAGGCCCGCGACGCCGTAAAGGCGTTTTGGGGCAACCGTGAAGCTGCACGGCAA
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

LYS GLN ILE GLU SER GLY LYS ALA ASP GLN GLY GLU ARG ALA GLY VAL THR GLY GLY LYS ASN MET ASP GLY PHE LEU ALA LEU VAL LEU ASP VAL ILE LYS ALA ASN GLY LEU ALA HIS
AAGCAGATTGAGTCAGGCAAGGCCGACCAAGGTGAACGCGCTGGCGTCACAGGCGGCAAGAACATGGACGGATTCTTGGCCTTGGTGCTCGATGTCATCAAGGCCAATGGTCTGGCCCAC
1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040

ALA GLU ILE HIS GLN ASN ARG ALA MET LEU THR LEU PRO GLY TYR PHE ARG PRO THR LYS LEU TRP ASP LEU LEU VAL ILE TYR LYS GLY GLU LEU ILE ALA ALA ILE GLU LEU LYS SER
GCCGAGATTCACCAGAACCGGGCGATGCTGACCCTGCCCGGCTACTTCCGCCCGACAAAGCTGTGGGATTTGTTGGTGATCTACAAGGGCGAGCTGATCGCAGCCATCGAACTGAAAAGC
2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160

```
HISVALGLYPROSERPHESERASNASNPHEASNASNARGTHRGLUGLUALAILEGLYTHRALAHISASPLEUTRPTHRALATYRARGGLUGLUALAPHEGLYLYSGLNPROARGPROPHE
CATGTCGGGCCATCGTTCAGCAACAATTTCAACAACCGTACCGAGGAAGCCATCGGAACAGCTCATGACCTCTGGACAGCCTACCGGGAAGAGGCTTTCGGCAAGCAGCCACGTCCCTTC
       2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

VALGLYTRPLEUMETMETVALGLUASPALAPROGLUSERARGARGPROVALARGASPSERSERPROHISPHEPROVALPHEGLUGLUPHELYSGLYALASERTYRLEUTHRARGTYRASP
GTCGGCTGGCTGATGATGGTCGAAGATGCGCCGGAGTCCCGCCGTCCCGTTCGGGACTCGTCGCCGCATTTCCCGGTCTTCGAGGAGTTCAAGGGGGCATCGTATCTCACGCGGTATGAC
       2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

LEULEUCYSGLNARGLEUVALGLNGLUGLNLEUTYRTHRTHRALAALAVALILEALAALAGLUARGSERALAVALASPTHRGLYASNPHETHRGLULEUSERSERMETTHRSERLEULYS
TTGCTGTGTCAGCGATTGGTGCAGGAGCAGCTCTATACGACTGCTGCCGTCATCGCGGCAGAACGCAGTGCGGTGGATACCGGCAACTTCACGGAGCTGTCGTCGATGACCAGCCTCAAG
       2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520

THRPHEVALSERALALEUALAGLYHISILEALAALAGLUALAALAARGLEUGLY
ACATTCGTGTCGGCTTTGGCTGGGCACATCGCCGCCGAGGCGGCACGACTGGGCTGATTGCCCGTTCTCGATCATTCGGGCTTGGAGGCCGGACGATGGTGACAACCACAACTGAGGCGC
       2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640

TTTCAGAGGTCTTGCGCGGGTATGGACGCGCTGATGAAACGGCATTGGCATCACTGCTGGAGGTCAGCGACTGGGAGCGCATCGCTCGTCAGGAGCTGGAGCGCCGGGCTACCCGAATTG
       2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760

TGCAGTCGCTTGATGACGGCACCCTTGAGGCCATCGCATCTGGTTCGCTCGATGTCCGGGCCGTCTGCCGCCAAGTGGCCGACGAGATCAAGCAGACCGCTTAGACGATGGCGCTGCCTT
       2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880

CTTTCCGACCAACGGGTGGCGCGAGGGCCAGAGCGTAAAGCCGCGTTACCTCGTGCTGGAGATCAAAGACAAGGATCGCCCGGATGACAAGGCATTGGGATGGGTGCTTGTCGAGCGGGA
       2890      2900      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000

GGAAACCTATCGGCGCGATCCTCGGGATGGGACGATCTACGAGGCGTCCATTCGGCTTTCCTACCAGCGCATCACAGCCAAGTTTTCGTACCGGGATGGTGGCAACAGGGCGATTTGATG
       3010      3020      3030      3040      3050      3060      3070      3080      3090      3100      3110      3120

GCAGCTACTCGCGCAACTTCAACTCTGTCTCGTTGACTTCCACCTCGATGTCGAAAGGGGCGTATTCCTTGATTTGCCGGGGCTGGATGGCCAGCGAATCGGCACCTACTTGATGAATGA
       3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240

GATCGTCCAGTGGGTGCAGCAGTGGCCCGAGACGACCGTCAATAGCATCGAGCTGCTGGCCGGACAGGCGCATGGTGACAACAAGGCACGGCGCAACTGGTTTTATGAGCAGTTCGGTCT
       3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360

GTCTTCGACTACACCGACCCCGAGGCATCGCGA
       3370      3380      3390
```

Figure 3. NUCLEOTIDE SEQUENCE OF *Pae*R7 RESTRICTION-MODIFICATION SYSTEM DERIVED FROM pPAORM3.8. Putative regulatory regions are labeled with "...." for the -10, and with "***" for the -35 regions. Two tandem *Pae*R7 recognition sites (CTCGAGGCTCGAG) are also highlighted. Also denoted (---) is the amino acid sequence of the endonuclease which was confirmed by direct protein sequencing. The first nucleotide of each of the *Bal*31 mutants, pPAOΔ594, pPAOΔ614 and pPAOΔ628, is marked. The *in vivo* phenotypes of these mutants are summarized in Figure 2.

*Nru*I sites by use of *Bal*31 nuclease digestions. Deletions from the *Nru*I site produced mutations in the structural region of first the endonuclease and then the methylase genes. Figure 2 presents the location of four of these structural gene mutants. Plasmid pPAORM3.2 is the smallest characterized plasmid capable of transferring (*in vivo* and *in vitro*) both restriction and modification phenotypes (R+M+) to *E. coli*. However, pPAO2.2 and pPAO1.6 are plasmids incapable of conferring either phenotype (R-M-).

*Bal*31 deletion mutants made from the *Bam*HI site 5' to the methylase gene demarcates the putative promoter region of this

system. The clone pPAOΔ594, the least-deleted of this series, conferred the same phenotype as pPAORM3.8 (R+M+) and had comparable levels of endonuclease and methylase activities in extracts. pPAOΔ614 is 20 base pairs shorter than pPAOΔ594. Although its in vitro activities and phage modification levels are indistinguishable from pPAORM3.8, the clones restrict Ø80 at least 10x less efficiently. pPAOΔ628, an additional 14 base pairs shorter, has even more unusual properties. This clone no longer detectably restricts Ø80 phage and only partially modifies progeny phage. There is no discernible methylase activity in crude extracts, and restriction activity is reduced over 1000x (i.e., equivalent to the level of pPAOR1.9) with pPAOΔ628. Both proteins are still being made but at a much lower level. These results are consistent with the region 5' to the methylase gene acting in a regulatory capacity, and with the PaeR7 system functioning as an operon. Since the endonuclease continues to be made even when this region and part of the methylase is deleted (as in pPAOR1.9), a secondary promoter may possibly be present within the methylase gene controlling endonuclease expression.

Both the chemical cleavage and dideoxynucleotide chain terminator methods for DNA sequence analysis were used to determine 3393 base pairs encoding the primary structures of the endonuclease and methylase genes within the 3800 base pair insert of pPAORM3.8 (Figure 3). In addition, the nucleotide sequences of several of the Bal31 deletion mutants were determined for the regions which constituted the junction sites between the vector and insert. Overlapping sequences were identified by computer analysis (29).

The PaeR7 sequence was analyzed for open reading frames beginning with an ATG codon and capable of encoding polypeptides of at least 150 amino acids in length (Figure 4). Two contiguous open reading frames were detected which could encode the endonuclease and methylase genes as a single operon. Such a gene organization is suggested by the Bal31 deletion mutants. The open reading frame for the putative methylase gene begins with an ATG at nucleotide 112 and terminates at 1834 (TAA) (Figure 3). Contiguous with this first open reading frame is a second open reading frame which could encode the endonuclease

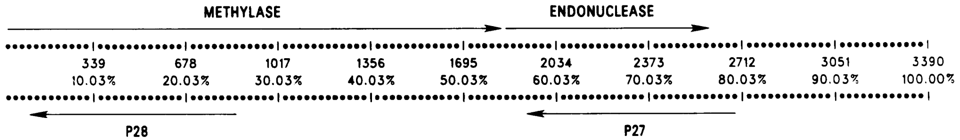


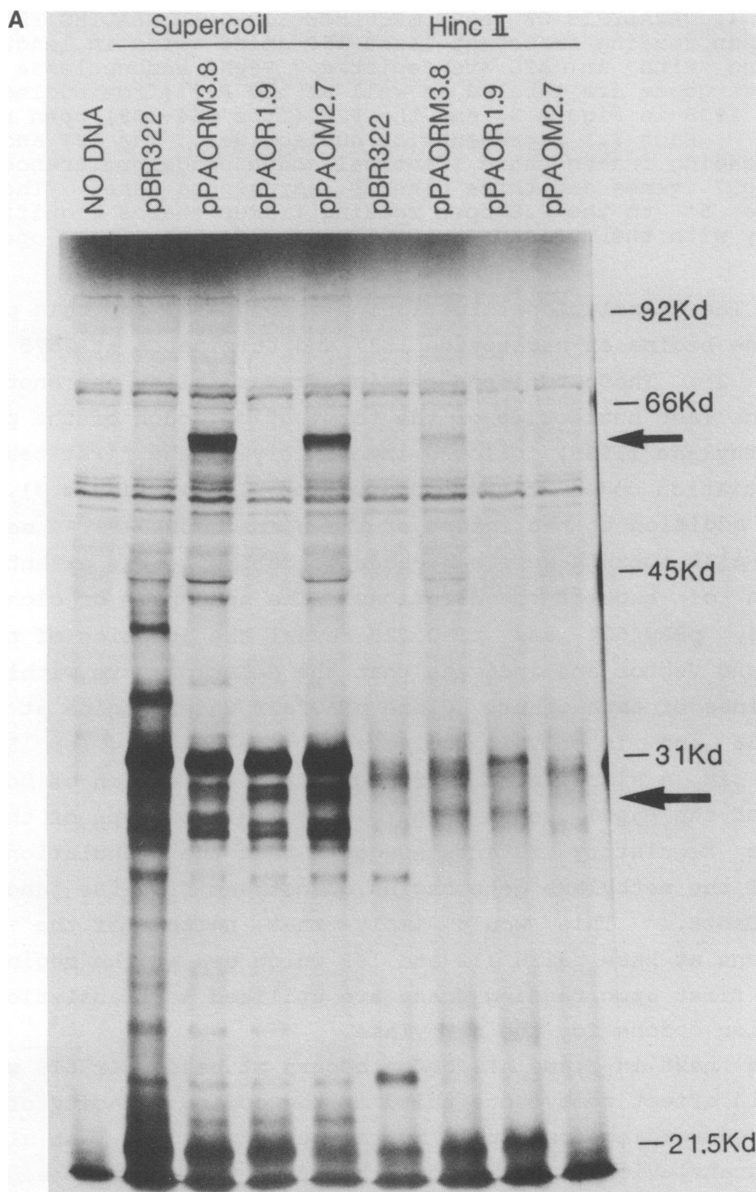
Figure 4. ANALYSIS OF PaeR7 SEQUENCE FOR OPEN READING FRAMES. Only open reading frames at least 150 amino acids in length and beginning with an ATG are depicted. PaeR7 endonuclease and methylase genes are labeled as well as the P27 (from nucleotides 2691 to 1935 in Figure 3) and the P28 (from 864-108) open reading frames. Each (.) represents 26 nucleotides. The P27 and P28 open reading frames share identical codon usage preference with the PaeR7 genes and three other P. aeruginosa genes. The DNA sequence 5' to the P28 open reading frames shares significant homology with the putative control region for the PaeR7 operon.

gene. The translational initiation codon (ATG) for this putative gene begins at nucleotide 1837 and terminates at 2575 (TGA) (Figure 3). These two open reading frames follow one another with the last nucleotide of the termination codon of the putative methylase (1836) followed immediately by the first base of the initiation codon of the endonuclease (1837) (Figure 3).

In addition to the insert of the plasmid pPAORM3.8, several of the Bal31 mutants were sequenced to determine the extent and location of each of the deletions. The sequences of clones pPAOΔ594, pPAOΔ614 and pPAOΔ628 reveal the junction of the insert and vector and indicate that the deletions are within 34 nucleotides of each other. Clone pPAOΔ628 which begins at base pair 166 and is only 14 nucleotides from pPAOΔ614 (bp 152) results in a significant decrease in the expression of both genes of the PaeR7 system (Figure 2). The phenotype of these putative regulatory mutants suggests that the translational start of the methylase gene begins downstream from the junction of pPAOΔ628. This would imply that neither of the ATG codons at base pairs 112 and 169 which are at the beginning of the first open reading frame are utilized as translational initiation codons for the methylase.

The next in-frame ATG codon occurs at base pair 670 which would in effect remove one third of the coding capacity of the first open reading frame used by the methylase gene. An alternative translational initiation site after the junction sequence of pPAOΔ628 is a GTG codon at base pair 241 which is the first

occurrence of such a codon after base pair 166 (pPAOΔ628). A methylase gene initiating at nucleotide 241 (GTG) would be 531 amino acids and have a predicted size of 59,260 daltons.



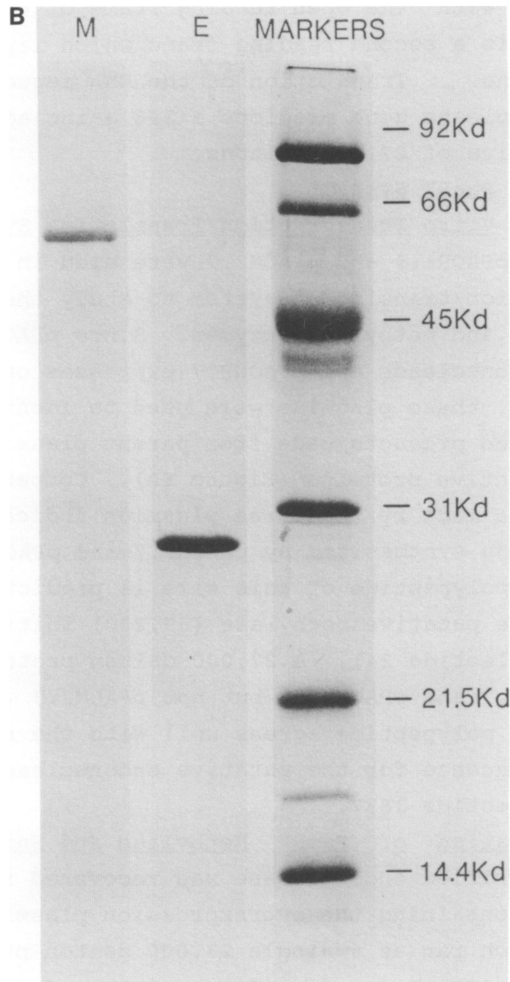


Figure 5A. IN VITRO TRANSCRIPTION-TRANSLATION PRODUCTS. An autoradiograph of 11% polyacrylamide gel containing methylase and endonuclease products synthesized from a coupled *in vitro* system. Plasmids were supercoiled plasmids as used in the coupled reaction. A 60,000 dalton protein (M) appears in lanes from pPAORM3.8 and pPAOM2.7 but not in the others. Size markers (Biorad) indicate the relative size of both proteins. As indicated by the manufacturer, the efficiency of transcription decreases using the *Hinc*II-linearized DNA in place of the supercoiled forms. *Hinc*II-cleaved forms of the plasmids are used to block the expression of β-lactamase protein in order to visualize the 27,000 dalton endonculease band.

B. SAMPLES OF PURIFIED *Pae*R7 METHYLASE AND ENDONUCLEASE RUN ON 15% SDS ACRYLAMIDE GEL WITH A 5% STACKING GEL. E = endonuclease and M = methylase.

Contiguous with the open reading frame of the putative methylase gene is a second reading frame which may encode the endonuclease gene. Translation of the DNA sequence of the putative endonuclease gene predicts a 246 amino acid polypeptide, having a size of 27,283 daltons.

Gene Products of PaeR7 System:

Coupled In Vitro Transcription/Translation Systems: Plasmids pPAORM3.8, pPAOR1.9 and pPAOM2.7 were used in a coupled *in vitro* transcription/translation system to study the synthesis of the endonuclease and methylase enzymes. Since pPAOR1.9 expresses only the endonuclease and pPAOM2.7 expresses only the methylase (Figure 2), these plasmids were used to identify which *in vitro* synthesized products made from parent plasmid pPAORM3.8 were the respective proteins (Figure 5A). Comparison of the protein products made by the three plasmids indicate a 60,000 dalton polypeptide synthesized by pPAOM2.7 and pPAORM3.8 but not by pPAOR1.9. A polypeptide of this size is predicted by the DNA sequence for the putative methylase (59,260) initiating with a GTG codon at nucleotide 241. A 27,000 dalton protein is synthesized by pPAOR1.9 and pPAORM3.8 but not pPAOM2.7 (Figure 5A). The size of this polypeptide agrees well with the size predicted by the DNA sequence for the putative endonuclease (27,283) starting at nucleotide 1837.

Characterization of PaeR7 Methylase and Endonuclease: Twelve mg of purified endonuclease was recovered from 41 g of *E. coli* cells containing the overexpression plasmid (Table 1). The enzyme, which ran as a single 27,000 dalton protein during SDS-PAGE (Figure 5B), has a specific activity of 6 x 10^6 units/mg. The specific activity of this purified *Pae*R7 endonuclease when measured on lambda DNA is 15 x 10^6 units/mg and is comparable to values found for homogenous *Bam*HI, *Pst*I and *Eco*RI endonucleases (J. Benner and L. Sznyter, New England BioLabs, unpublished observations). In contrast, we recovered 0.4 mg of purified methylase (30x less than endonuclease) from these cells. The methylase ran as a single band of 60,000 daltons on SDS-PAGE (Figure 5B). It had a specific activity of 10^5 units/mg.

The active forms of the enzymes were sized by gel filtration. The endonuclease and methylase activities eluted with estimated sizes of 52,000 and 58,000 daltons, respectively.

This is consistent with the endonuclease forming a dimer and the methylase acting as a monomer.

The amino acid sequence of the first 20 residues of the purified *Pae*R7 endonuclease was determined by Edman degradation. Assuming post-translational removal of the first amino acid, this sequence agrees precisely with that predicted by the DNA sequence (Figure 3). Attempts to sequence the methylase by the same procedure proved unsuccessful.

DISCUSSION

Restriction-modification systems differ markedly in their organization. The only common feature thus far found among the characterized systems is that the two genes are closely linked. The organization of the *Pae*R7 system most closely resembles the chromosomal *Hha*II system in that the genes are cotranscribed in the order methylase to endonuclease (5). In the best-studied system, *Eco*RI, the order is reversed with the endonuclease preceding the methylase on a common transcription unit (8,9). In three other systems characterized, *Pst*I (7), *Eco*RV (11) and *Pvu*II (12), the two promoters for the endonuclease and methylase are situated in an intergenic region and are divergently transcribed. From the diversity of these gene arrangements, it would seem that the mode of coordinate control between the endonuclease and methylase genes must vary.

The regulatory region for the *Pae*R7 operon has been operationally defined by characterizing several *Bal*31 deletion mutants. The DNA sequences of these mutants provide supporting data which indicate that deletions within a 34 nucleotide region, 5' to the putative start of the methylase gene, have pleiotropic effects. Plasmids pPAOΔ594, pPAOΔ614 and pPAOΔ628 demonstrate a spectrum of decreasing levels of expression for both the proximal methylase and distal endonuclease genes. Inspection of the nucleotide sequence of these *Bal*31 deletion mutants reveals that beginning at nucleotide 165 (76 base pairs away from the putative translational start for the methylase gene) the sequence TGTCATG occurs. This sequence has been reported to be the Pribnow region (-10 region) for the *E. coli* *ara*C gene (30). Although this observation is disputed by RNA transcription mapping experiments of this gene (31), there is

evidence that indicates this sequence may be capable of acting as a secondary promoter for the araC gene (G. Wilcox, personal communication). The sequence TTGCCG beginning at nucleotide 146, has been designated as the -35 region of the PaeR7 promoter. Although closer to the Pribnow region than is usually found in E. coli promoters [13 base pairs instead of 15-21 (32)], this -35 sequence shows the most resemblance to the consensus -35 region of E. coli promoters (TTGACA) (33). It is uncertain whether the sequence we have identified as the putative control region for the PaeR7 system operates as such in P. aeruginosa.

The molar amount of endonuclease purified exceeded that of methylase by 30-fold (data not shown). It is unlikely that the discrepancy resulted from differential losses during purification. ^{35}S-labeled cell extracts run on 2-D polyacrylamide gels show greater than 10x excess of endonuclease in crude extracts (data not shown). It is curious that the yield of endonuclease and methylase vary so greatly since the two genes are presumably cotranscribed with the methylase preceding the endonuclease. If both genes are being translated from the message with equal efficiency one would expect more methylase than endonuclease in the cell since: a) the methylase polypeptide is twice the size of the endonuclease and b) the methylase is active as a monomer and the endonuclease as a dimer.

This suggests that there is a greater turnover of methylase than endonuclease within the cell and/or that the endonuclease is translated more efficiently than the methylase. Comparison of putative ribosome binding sites and translation initiation codons favors the latter possibility. In the region proximal to the putative start of the methylase there is no occurrence of a sequence resembling the consensus Shine-Dalgarno [i.e. GGAGG; (34)]; however, nine bases 5' to the initiation of the endonuclease gene is the sequence GGAGA. It may be significant that other methylase genes that have been sequenced (i.e., HhaII [5], PstI [7], EcoRV [11], Bacillus subtilis phage SPR [35] and dam [36]) also have poor ribosome binding sites. In addition, it is possible that use of the infrequent initiation codon GTG rather than the more conventional ATG reduces methylase expression.

The translational initiation codon for the endonuclease gene is an ATG starting at nucleotide 1837 (Figure 3); the initiating fMET of this polypeptide is removed in E. coli. Similar processing has been observed for the EcoRI endonuclease and methylase (8), as well as for the PstI methylase (7) and the EcoRV endonuclease (11). However, such processing is not universal for all endonuclease-methylase products; PstI endonuclease (7) and EcoRV methylase retain their fMET residues (11). Other types of protein processing are missing from the PaeR7 products. The sequence data for both the PaeR7 methylase and endonuclease genes confirm that neither enzyme possesses a hydrophobic signal peptide sequence at the amino-termini (Figure 3). The protein products observed as a result of the coupled transcription-translation and the PaeR7 enzyme purifications support this observation. The lack of such signal peptides at the amino termini suggest that these proteins are not transported through a membrane system of the host by a canonical transport mechanism, as exemplified by the beta-lactamase enzyme (37).

Additional analysis of the PaeR7 sequence suggests other levels of control of expression. During translation, there is in bacteria a non-random use of codons which is reflected in the choice of the nucleotide at the third position (38,39). Comparison of the codon usage of the PaeR7 genes to the strongly and weakly expressed genes of E. coli demonstrates a similarity to the strongly expressed E. coli genes (40). Thus, the codon usage for the PaeR7 genes probably contributes to the reasonable levels of expression of these genes in E. coli. The PaeR7 methylase (61.5%) and endonuclease (59.3%) have a combined G+C content of 60.8%. Comparison of the codon usage of these PaeR7 genes and three other Pseudomonas genes (41,42,43) show the same preference for G or C nucleotides in the third position of a codon, thus resulting in identical codon preferences except for the amino acids Val, Thr, Asn, Asp, Cys and Ser.

Finally, there appears to be control of bacteriophage restriction as a result of post-translational events. There is evidence from complementation experiments using subclones of the PaeR7 system which suggests that both parts of the system are

required for efficient phage restriction. The details of this observation are covered in a companion paper (17).

ACKNOWLEDGEMENTS

We would like to thank J. Garrels for his assistance in analyzing the purified endonuclease and methylase protein by two-dimensional electrophoresis and K. Silber for her technical assistance. The plasmid pGW7 was a gift from G. Wilson (New England BioLabs). We thank G. Davis, D. Kwoh and J. Kwoh for careful reading and discussion of this paper. We also thank Susan Konchal and Janice Doty for their help in the preparation of this manuscript. This research is supported by NSF (PCM84-02963) and SIBIA grants to T.R.G.; NSERC (A-6774) and FCAR (EQ-1700) grants to P.H.R. and by NSF (PCM 8217553) to R. J. Roberts while some of this work was being carried out at Cold Spring Harbor Laboratory.

REFERENCES

1. Yuan, R. (1981) Ann. Rev. Biochem. 50, 285-315.
2. Modrich, P. and Roberts, R. J. (1982) in Nucleases, Linn, S. M. and Roberts, R. J. Eds. pp. 109-154. Cold Spring Harbor Laboratory, New York.
3. Bickle, T. A. (1982) in Nucleases, Linn, S. M. and Roberts, R. J. Eds. pp. 85-108. Cold Spring Harbor Laboratory, New York.
4. Mann, M. B., Rao, R. N. and Smith. H. O. (1978) Gene 3, 97-112.
5. Schoner, B., Kelly, S. and Smith, H. O. (1983) Gene 24, 227-236.
6. Walder, R. Y., Hartley, J. L., Donelson, J. E. and Walder, J. A. (1981) Proc. Natl. Acad. Sci. USA 78, 1503-1507.
7. Walder, R. Y., Walder, J. A. and Donelson, J. E. (1984) J. Biol. Chem. 259, 8015-8026.
8. Newman, A. K., Rubin, R. A., Kim, S. H. and Modrich, P. (1981) J. Biol. Chem. 256, 2131-2137.
9. Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E. and Rosenberg, J. M. (1981) J. Biol. Chem. 256, 2143-2152.
10. Kosykh, V. G., Buryanov, Y. I. and Bayev, A. A. (1980) Mol. Gen. Genet. 178, 717-719.
11. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) Nucl. Acid. Res. 12, 3659-3676.
12. Blumenthal, R. M., Gregory, S. A. and Cooperider, J. S. (1985) J. Bacteriol. (in press).
13. Theriault, G. and Roy, P. H. (1982) Gene 19, 355-359.
14. Gingeras, T. R. and Brooks, J. E. (1983) Proc. Natl. Acad. Sci. USA 80, 402-406.
15. Jacoby, G. A. and Sutton, L. (1977) Plasmid 1, 115-116.
16. Hinkle, N. and Miller, R. V. (1979) Plasmid 2, 387-393.

17. Brooks, J.E., Gingeras, T.R., and Howard, K. (1985) Gene (in press).
18. Clewell, D. B. and Helinski, D. R. (1972) J. Bacteriol. 110, 1135-1146.
19. Birnboim, H. C. and Doly, J. (1979) Nuc. Acid Res. 7, 1513-1523.
20. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
21. Messing, J., Crea, R. and Seeburg, P. H. (1981) Nucl. Acids. Res. 9, 309-321.
22. Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101-106.
23. Cohen, S. N., Chang, A. C. Y., and Hsu, L. (1973) Proc. Natl. Acad. Sci. USA 69, 2110-2114.
24. Maxam, A. M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds. Vol. 65, pp. 449-560, Academic Press, New York.
25. Sanger, F., Nicklen S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
26. Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3967.
27. Laemmli, U. K. (1970) Nature (London) 227, 680-685.
28. Bradford, M. M. (1976) Analyt. Biochem. 72, 248-254.
29. Gingeras, T. R., Milazzo, J., Sciaky, D., Roberts, R. J. (1979) Nucl. Acid Res. 7, 529-545.
30. Smith, B. R. and Schleif, R. (1978) J. Biol. Chem. 253, 6931-6933.
31. Wallace, R. G., Lee, N., and Fowler, A. V. (1980) Gene 12, 179-190.
32. Hawley, D. K. and McClure, W. R. (1983) Nucleic Acids Research 11, 2237-2255.
33. Siebenlist, U. (1979) Nucl. Acids Res. 6, 1895-1907.
34. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
35. Buhk, H.-T., Behrens, B., Tailor, R., Wilkie, K., Prada, J. J., Gunthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T. A. (1984) Gene 29, 51-61.
36. Brooks, J. E., Blumenthal, R. M. and Gingeras, T. R. (1983) Nucl. Acid Res. 11, 837-851.
37. Sutcliffe, J. G. (1978) Proc. Natl. Acad. Sci. USA 75, 3737-3741.
38. Wain-Hobson, S., Nussinov, R., Brown, R.J. and Sussman, J.L. (1981) Gene 13, 355-364.
39. Bibb, N.J., Findlay, P.R. and Johnson, M.W. (1984) Gene 30, 157-166.
40. Grosjean, J. and Fiers, W. (1982) Gene 18, 199-209.
41. Minton, N.P., Atkinson, T., Bruton, C.J. and Sherwood, R.F. (1984) Gene 31, 31-38.
42. Brown, N.L., Ford, S.J., Pridmore, R.D. and Fitzinger, D.C. (1983) Biochemistry 22, 4089-4095.
43. Nakai, C., Kagamiyama, H., Nozaki, M., Nakazawa, T., Inouye, S., Ebina, Y., and Nakazawa, A. (1983) J. Biol. Chem. 258, 2923-2928.