Vol. 107, No. 2, 1982 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 30, 1982 Pages 568-575

NUCLEOTIDE SEQUENCE OF THE PROMOTER REGION OF THE GENE CLUSTER FOR PROTON-TRANSLOCATING ATPASE FROM ESCHERICHIA COLI AND IDENTIFICATION OF THE ACTIVE PROMOTER

Hiroshi Kanazawa, Kazunori Mabuchi and Masamitsu Futai

Department of Microbiology, Faculty of Pharmaceutical Sciences, Okayama University, Okayama 700, Japan.

Received June 15, 1982

Summary: A 505-nucleotide long DNA sequence of a part of the gene cluster for the proton-translocating ATPase (pap operon) of E.coli was determined. In the sequence determined upstream of the gene for 14K protein, two promoter sequences with reverse directions to each other were found. One of them was identified as the active promoter for the operon by in vitro transcription and the DNase I footprinting technique. The other was located upstream of the pap operon promoter and identified as an active promoter of a gene coding for a small RNA of unknown function adjacent to the operon.

The proton-translocating ATPase of E. coli catalyzes the hydrolysis and synthesis of ATP, reversibly. The enzyme is composed of eight independent subunits (α, β, γ, δ, ε, a, b, and c)(1,2). All the genes for the subunits are located in the 83 minute region of the E. coli linkage map and form a cluster as an operon (pap operon)(3-5,15). The nucleotide sequence of all the genes for subunits was determined using cloned genomic DNA (5-9) and in addition a novel gene was found coding for a Mr 14K protein (8). Thus the organization of the structural genes in the operon was definitely determined at the level of nucleotide sequence. However, no functional promoter of the pap operon has yet been identified.

In the present study, we determined the DNA sequence of 505 base pairs upstream from the amino terminal of the gene for the 14K protein and found a typical promoter sequence. In vitro transcription using a DNA segment containing this sequence showed that the sequence corresponds to a functionally active promoter for the operon. A possible initiation site(s) of transcription was also identified. Furthermore, a site for RNA polymerase-binding was determined by the DNase I footprinting technique (10). We also found another functional promoter sequence for a gene coding for a RNA of unknown function adjacent to the pap operon.

Abbreviations: 14K protein, a protein coded by a novel reading frame. (see text for details) ; Mr, molecular weight.

0006-291X/82/140568-08$01.00/0

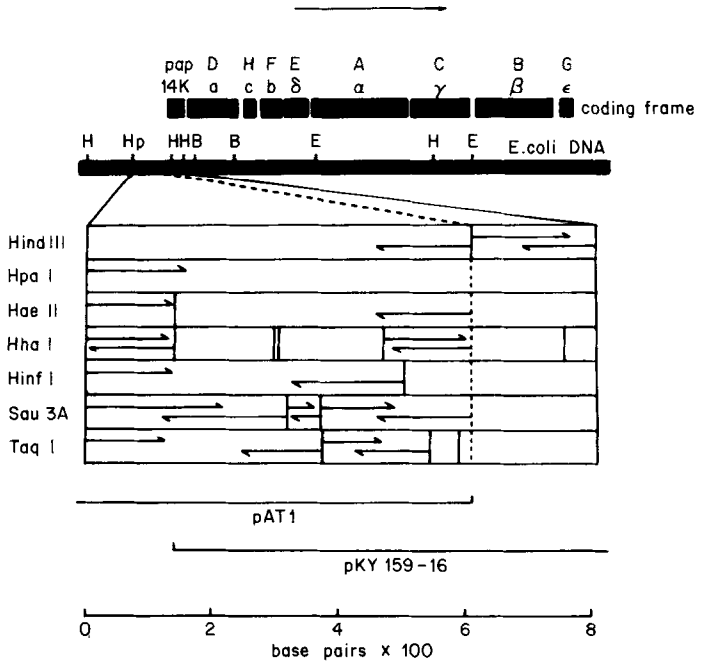


Fig. 1 Organization of the pap (unc) operon and DNA sequencing strategy. The direction of transcription of the gene cluster is shown at the top of the figure. The coding frame of each gene with its nomenclature (pap) (5) and coding subunit are shown above the E.coli DNA. Cleavage sites for endonucleases are as follows: E, EcoRI; H, HindIII; B, BamHI; Hp, HpaI. The cleavage maps with HaeII, HhaI, HinfI, Sau3A and TaqI are also shown. Arrows indicate the sequenced DNA segments with their directions and approximate lengths. Plasmids pAT1 and pKY159-16 cover the regions shown. The scale shown at the bottom corresponds to the numbers of nucleotide residues in Fig. 2.

MATERIALS AND METHODS

Preparation of plasmids and their segments: Hybrid plasmids pKY159-16 (8) and pAT1 (11) were used in the present study. The DNA segments used for sequencing, *in vitro* transcription and DNase I footprinting experiments were prepared by digesting plasmid DNA with various restriction endonucleases (Takara Shuzo Co., Japan).

Determination of the DNA sequence: DNA fragments prepared by the sequencing strategy (Fig. 1) were phosphorylated at the 5' end with ^{32}P-γ-ATP and T4-polynucleotide kinase (Boehringer-Mannheim). The DNA sequence was determined by the method of Maxam and Gilbert (12).

***In vitro* transcription:** Various DNA segments derived from the promoter region were transcribed *in vitro* with RNA polymerase from E. coli by the procedure described previously (13). About 3 μg of DNA and 1 μg of purified E.coli RNA polymerase holoenzyme (New England Biolabs.) were mixed in 50 μl of reaction mixture consisting of 40 mM Tris-HCl, pH 7.9, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 50 mM KCl, 12.5 % glycerol, 0.5 mM ATP, 0.5 mM UTP, 50 μM GTP, 0.39 μM ^{32}P-α-GTP (Amersham) and incubated for 2 hr at 37 C. The reaction was terminated by heating the reaction mixture for 3 min. The sizes of the RNAs thus synthesized were analyzed by polyacrylamide gel electrophoresis (8 % w/v acrylamide, 8 M urea) and subsequent autoradiography.

DNase I footprinting: The experimental procedure was that of Galas and Schmitz (10). A DNA segment was phosphorylated at the 5' end with ^{32}P-γ-ATP and T4-polynucleotide kinase. About 3 μg of a phosphorylated DNA segment was digested with various amounts of pancreatic DNase I

(Worthington) in 50 µl of solution (20 mM Tris-HCl, pH 7.9, 50 mM KCl, 10 mM $CaCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 5 % glycerol) in the presence of 13 µg of purified RNA polymerase for 30 sec at 25 C. Then 150 µl of a solution of 0.3 M sodium acetate, containing 0.1 M EDTA and 50 µg/ml tRNA was added to stop the reaction. The mixture of DNA fragments thus prepared was subjected to polyacrylamide gel electrophoresis and autoradiography.

RESULTS AND DISCUSSION

DNA sequence of the promoter region for the pap operon:

We determined a 505-nucleotide long sequence of DNA adjacent to the amino terminal portion of the gene for the 14K protein of the pap operon (Fig. 2). Two sequences similar to the consensus sequences of the "Pribnow box" and -35 region (a recognition site of RNA polymerase) (14) were found at positions about 250 (E + D) and 480 (A + B) in Fig. 2. The one (A+B) closer to the gene for the 14K protein, has the same direction as that for transcription of the pap operon and was identified as an active promoter, as described below.

In vitro transcription:

To determine whether the two promoters found in the DNA sequence are active, a DNA segment carrying both sequences was subjected to in vitro transcription by a purified RNA polymerase and the sizes of the transcripts were analyzed. (Fig. 3). Two major transcripts, comprised of about 120 (transcript A) and 85 (transcript B) nucleotides, were found using a template of 615 nucleotide long DNA (b segment), carrying the two candidates for the promoter site and a part of the gene for the 14K protein (Fig. 3, b). These results indicate that two active promoters exist in the DNA segment, because the transcripts were shorter than the product expected for non-specific read-through of the template. Furthermore, the coding regions for the respective transcripts were estimated by analyzing the sizes of transcripts using DNA segments carrying different portions of the b segment (Fig. 3, a, c, and d segments). The a and c segments coded for the transcripts B and A, respectively (Fig. 3), indicating that both segments have promoter sites. When the d segment, about 57 nucleotide shorter than the c segment, was used for transcription, approximately 57 nucleotide shorter RNA (transcript C) was synthesized (Fig. 3, d). This result suggests that an active promoter exists in the d segment and the direction of transcription is the same as that of the pap operon estimated previously (Fig. 3) (1, 15).

The 5' end of mRNA in E.coli is adenine (A) or guanine (G) residue in most cases determined so far and is located about four to seven nucleotides downstream from the end of the "Pribnow box" (14). Therefore, we concluded that the initiation site of transcript A is possibly G (position 492, Fig. 2). If this is the case, the transcript should be 127

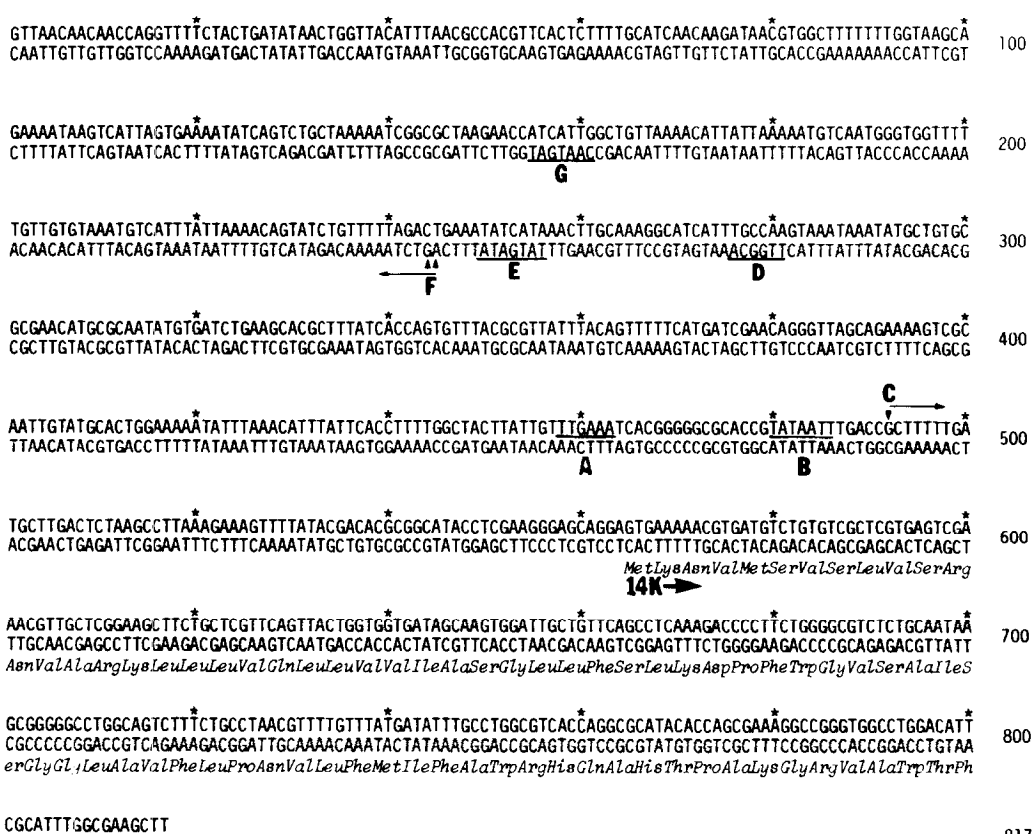
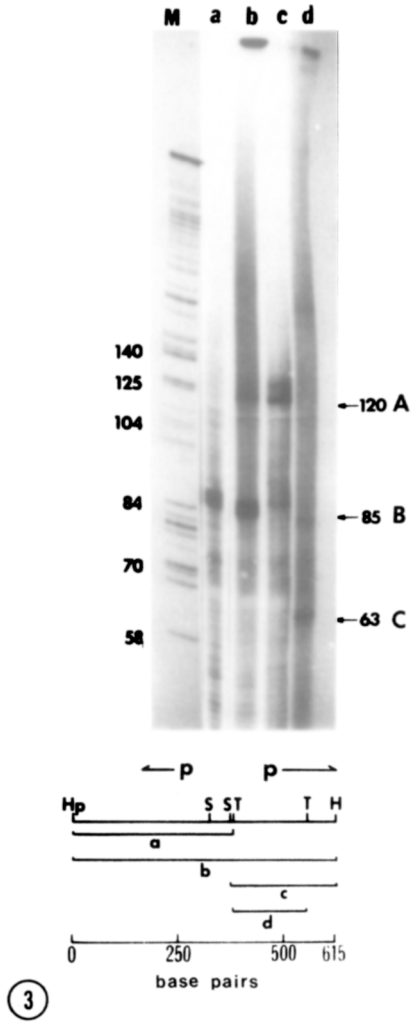


Fig. 2 DNA sequence of the promoter of the pap operon and flanking region. The DNA sequence in the antisense strand is shown with the deduced amino acid sequence. The numbers of residues shown on the right correspond to the scale in Fig. 1. A and D, a sequence recognized by RNA polymerase (-35 region) ; B and E, the "Pribnow box" ; C, (▼) a residue corresponding to the 5' end of mRNA of the pap operon. The arrow indicates the direction of transcription ; F, a residue corresponding to the 5' end of the newly identified RNA of unknown function. ; G, a possible terminator sequence for the RNA. This sequence is similar to that found in several genes for which termination is dependent on the rho factor. The DNA sequence after position 389 in Fig. 2 was determined independently by Gay and Walker by the dideoxynucleotide method (19). Our sequence agrees with theirs. We reported the DNA sequence after position 506 previously(8), but several nucleotide residues between position 560 and 817 were revised in the present study.

and 62 nucleotide long with the c and d segments respectively, as templates. These numbers agreed well with those found in in vitro transcripts (120 and 63 nucleotide long). The difference between the values for the sizes of transcripts calculated and synthesized in vitro may be due to a slight difference between migrations of the single stranded DNA used as a marker (Fig. 3, M) (a sense strand) and that of transcribed RNA (an antisense strand). Thus G (position 492) is concluded to be the initiation site of transcript A.



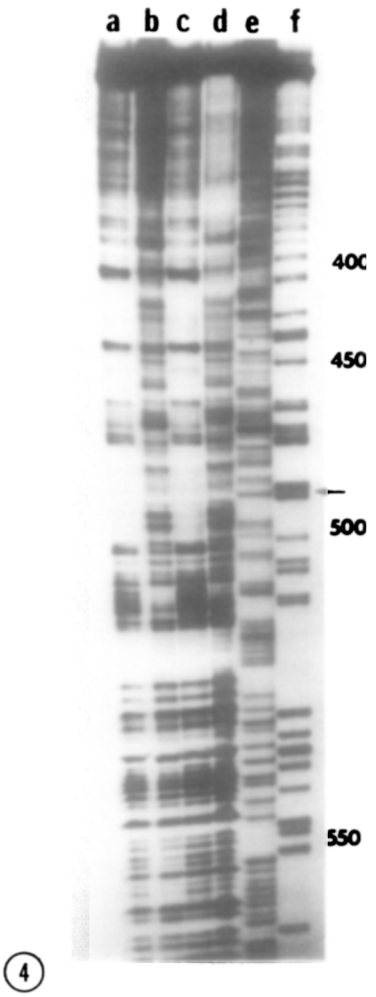


Fig. 3 In vitro transcription of various DNA segments carrying promoter regions. Various portions (a-d) of the DNA segment carrying the promoter regions used for in vitro transcription are shown at the bottom of the figure. The cleavage sites correspond to those in Fig. 1. P and arrows indicate the promoter site and the direction of transcription, respectively. The same amount of each DNA segment (adjusted by molecular weight of each segment, about 3 μg) was incubated with a RNA polymerase as described in the Materials and Methods. Samples were subjected to polyacrylamide gel electrophoresis (8 % (w/v) acrylamide, 8 M urea) and autoradiography. The letters (a-d) at the top of the figure correspond to the DNA template shown at the bottom. M indicates a size-marker. The b segment was labeled at the 5' end with ^{32}P-γ-ATP and T4-polynucleotide kinase and was digested with restriction endonuclease HaeII. Half the segment with the HindIII site was recovered after gel electrophoresis. This segment was subjected to chemical cleavage specific to the G residue and applied on gel as a size-marker. Numbers on the left and numbers at the right indicate the G residues from the HindIII site (Fig. 1) and the approximate sizes of major transcripts (A, B and C) based on the marker, respectively. One or two minor, but relatively strong bands adjacent to each major transcript (A, B and C) were observed. These bands may have appeared because termination in each case is not specific.

Fig. 4 DNase I footprinting analysis. A DNA segment extending from the HpaI site (position 1) to the HindIII site (position 615) was labeled at the 5' end and digested with HaeII. Half the segment with HindIII site was

A possible initiation site of transcript B is G (position 284) or A (position 285). Although we did not analyze the direction of transcription, a typical terminator sequence was found in a region around position 160 (G region) (Fig. 2). This region was consistent with that expected from the size of the transcript (about 85 nucleotides, Fig. 3, a). The terminator sequence is similar to those found in genes in which termination depends on the rho factor (14). Thus we believe that transcript B is coded by a DNA segment extending from about position 160 to 244 (or 245).

Binding site of RNA polymerase in the promoter for the pap operon:

To determine the binding site of RNA polymerase in the promoter of the pap operon and confirm that the promoter is active, we determined sequence of the DNA protected from attack by DNase in the presence of RNA polymerase (DNase I footprint analysis). The intensities of residues (Fig. 4) in sequences including the "Pribnow box" (about position 480) and the -35 region (about position 460) were much lower in the presence of RNA polymerase than in its absence, suggesting that these residues were protected from attack by DNase I (Fig. 4). The results of DNase I footprint analysis are summarized in Fig. 5. The protected region, possibly the binding site of RNA polymerase, was from +12 to -84 residue, except for the six residues discussed below. It should be noted that the residues (positions -85 to -89 and +13) adjacent to the protected region and six residues (-15, -17, -27, -40, -49 and -50) within the protected region became more susceptible to DNase I on binding of RNA polymerase (Fig. 5). It was reported that the -27 residue in the β-lactamase promoter (16) and the -26 residue in the lac promoter (17) became more susceptible to DNase I on binding of RNA polymerase. These results indicate that residues around -27 are extruded to the outside of complex of the RNA polymerase and the DNA.

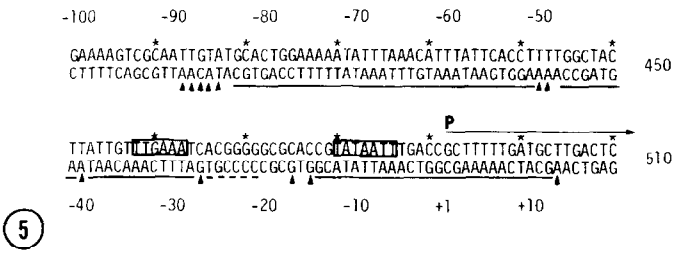
Promoter for the pap operon:

The results described above indicated that the sequence A and B (Fig. 2), located upstream from the gene for the 14K protein, is the real promoter for the pap operon, although the complete mRNA of the operon has not yet been determined. We found several similar sequences to the

---

Fig. 4 continued

ncubated for 30 sec at 25 C with various amount of DNase I (a and b, 2 µg/ml ; c and d, 4 µg/ml) in the presence (a and c) or absence (d and e) of RNA polymerase. Samples were subjected to polyacrylamide gel electrophoresis (15 % (w/v) acrylamide, 8 M urea) and autoradiography. The labeled DNA segment used for footprinting was cleaved at the specific site for C + T (e) or G (f) and applied to gel as a size marker. Numbers on the right correspond to residues in Fig. 2. The arrow indicates the suggested initiation site of transcription.



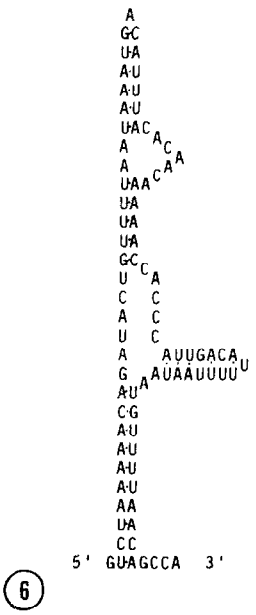


Fig. 5 Summary of DNase I footprinting experiments. The DNA sequence around the promoter of the pap operon protected from DNase I digestion in the presence of RNA polymerase is shown (underlined) with flanking sequences. Numbers above the sequence indicate residues, taking the first residue of mRNA as +1. Numbers on the right correspond to those in Fig. 1. The "Pribnow box" (TATAATT) and -35 region (TTGAAA) are shown in boxes. The arrow indicates the direction of transcription. Markers (▲) indicate enhanced digestion by DNase I. The broken line indicates the region of partial digestion.

Fig. 6 Possible secondary structure of a 86 nucleotide long RNA transcribed from a gene newly found in this study. The RNA shown is transcribed from a DNA segment (position 159-244). The G residue (position 244) is taken as the initiation site.

"Pribnow box" within the pap operon, but they do not necessarily have a corresponding -35 region (8). It has been suggested that the gene cluster for the proton-translocating ATPase is transcribed as a single unit, because Mu phage insertion had a strong polar effect (17). Therefore, we believe that sequences A and B are the unique promoter site for the operon. However, we cannot rule out the existence of a weak promoter(s), in which the sequence is apparently different from the consensus sequence in the operon. More detailed experiments on in vitro transcription using DNA segments from each portion of the pap operon are required for detection of a weak promoter(s).

A new gene coding for a small RNA:

As described above, we found a novel gene adjacent to the pap operon and identified its transcript. Although the RNA (86 nucleotides) is similar in size to tRNA, its secondary structure is different from the characteristic structure of tRNA (18) (Fig. 6). It has been reported that a gene for the

Mr 25,000 protein is located around this region (20). However, this RNA differs in size from the mRNA of the protein. It will be interesting to determine the real product and function of this newly identified gene.

Acknowledgements

We are grateful to Drs. T. Sekiya, M. Seiki, and H. Yoshikawa for suggestions on experimental procedures. This work was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan and Toray Science Foundation.

REFERENCES

1. Futai, M., and Kanazawa, H. (1980) Curr. Top. Bioenerg., 10, 181-215
2. Fillingame, R.H. (1981) Curr. Top. Bioeneg., 11, 35-106
3. Kanazawa, H., Miki, T., Tamura, F., Yura, T., and Futai, M. (1979) Proc. Natl. Acad. Sci. U.S.A., 76, 1126-1130
4. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T., and Futai, M. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 7005-7009
5. Kanazawa, H., Mabuchi, K., Kayano, T., Tamura, F., and Futai, M. (1981) Biochem. Biophys. Res. Commun., 100, 219-225
6. Mabuchi, K., Kanazawa, H., Kayano, T., and Futai, M. (1981) Biochem. Biophys. Res. Commun., 102, 172-179
7. Kanazawa, H., Kayano, T., Mabuchi, K., and Futai, M. (1981) Biochem. Biophys. Res. Commun., 103, 604-612
8. Kanazawa, H., Mabuchi, K., Kayano, T., Noumi, T., Sekiya, T., and Futai, M. (1981) Biochem. Biophys. Res. Commun., 103, 613-620
9. Kanazawa, H., Kayano, T., Kiyasu, T., and Futai, M. (1982) Biochem. Biophys. Res. Commun., 105, 1257-1264
10. Galas, D.J., and Schmitz, A. (1978) Nucleic Acid Res., 5, 3157-3170
11. Tamura, F., Kanazawa, H., Tsuchiya, T., and Futai, M. (1981) FEBS Lett., 127, 48-52
12. Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A., 74, 560-564
13. Kupper, H., Sekiya, T., Rosenberg, M., Eagan, J., and Landy, A. (1978) Nature, 273, 423-428
14. Rosenberg, M., and Court, D. (1979) Ann. Rev. Genet., 13, 319-353
15. Downie, J.A., Gibson, F., and Cox, G.B. (1979) Ann. Rev. Biochem., 48, 103-131
16. Russell, D.R., and Bennett, G.N. (1981) Nucleic Acid Res., 9, 2517-2533
17. Schmitz, A., and Galas, D.J. (1979) Nucleic Acid Res., 6, 111-137
18. Gibson, F., Downie, J.A., Cox, J.B., and Radick, J. (1978) J. Bacteriol., 134, 728-736
19. Gauss, D.H., Gruter, F., and Sprinzl, M. (1979) Nucleic Acid Res., 6, r1-r44
20. von Meyenberg, K., and Hansen, F.G. (1980) ICN-UCLA Symposia on Molecular and Cellular Biology, 19, 137-159
21. Gay, N.J., and Walker, J.A. (1981) Nucleic Acid Res., 9, 3919-3926