Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10380–10384, November 1992
Biochemistry

Identification of the gene for an *Escherichia coli* poly(A) polymerase

GONG-JIE CAO AND NILIMA SARKAR*

Department of Metabolic Regulation, Boston Biomedical Research Institute, Boston, MA 02114; and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Communicated by Charles C. Richardson, June 23, 1992 (received for review February 22, 1992)

ABSTRACT **Many bacterial mRNAs, like those of eukaryotes, carry a polyadenylate sequence at their 3′ termini, but neither the function of the bacterial poly(A) moieties nor their biosynthesis have been elucidated. To develop a genetic tool to approach the problem of bacterial poly(A) RNA, we have sought to identify the genes responsible for mRNA polyadenylylation. A poly(A) polymerase was purified to homogeneity from extracts of *Escherichia coli* and subjected to N-terminal sequence analysis. The 25-residue amino acid sequence obtained was used to design primers for the amplification of the corresponding coding region by the PCR from an *E. coli* DNA template. A 74-base-pair DNA segment was obtained that matched a region in the *pcnB* locus of *E. coli*, a gene that had originally been identified as controlling plasmid copy number [J. Lopilato, S. Bortner & J. Beckwith (1986) *Mol. Gen. Genet.* 205, 285–290] and was subsequently cloned and sequenced [J. Liu & J. S. Parkinson (1989) *J. Bacteriol.* 171, 1254–1261]. Direct evidence that the *pcnB* locus encodes poly(A) polymerase was provided by the observation that a bacterial strain transformed with an inducible expression vector carrying *pcnB* as a translational fusion produced 100-fold elevated levels of poly(A) polymerase upon induction. No increased poly(A) polymerase activity was observed in cells transformed with expression vectors carrying truncated forms of the *pcnB* gene. The identification of a gene encoding bacterial poly(A) polymerase opens the way for the study of the biosynthesis and function of bacterial polyadenylylated mRNA.**

Poly(A) polymerase catalyzes the template-independent sequential addition of AMP to the 3′-terminal hydroxyl groups of RNAs, in the following overall reaction: RNA + nATP → nPP$_i$ + RNA(A)$_n$.

In eukaryotes, poly(A) polymerase is essential for the maturation of mRNA, adding poly(A) tails to the 3′ end of precursor RNA generated by endonucleolytic cleavage (1). Neither the specific role of poly(A) tail nor its mechanism of formation is fully understood, but recent success in the cloning of the genes for bovine poly(A) polymerase (2, 3), vaccinia virus poly(A) polymerase (4), and yeast poly(A) polymerase (5) should facilitate the elucidation of the biological role of mRNA polyadenylylation. In prokaryotes, a poly(A) polymerase activity was partially purified from *Escherichia coli* (6), and enzyme preparations of various purities and different properties have been described (7–9), but these enzymes were not studied further because their function was obscure. However, the recent discovery of bacterial poly(A) RNA has again focused our attention on poly(A) polymerase in *E. coli* as the possible source of the poly(A) moieties at the 3′ ends of *E. coli* mRNA.

Studies from this laboratory showed that substantial amounts of poly(A) sequences exist in prokaryotic mRNA from *Bacillus brevis* (10), *Bacillus subtilis* (11), and *E. coli* (12). The presence of large amounts of poly(A) RNA in *B. subtilis* was confirmed by the oligo(dT)-dependent synthesis of cDNA by using reverse transcriptase (13) and made possible the construction of the first cDNA library from bacteria (14). Finally, two specific mRNAs in *E. coli*, namely those for the outer membrane lipoprotein (15) and tryptophan synthetase α subunit (16), were found to be extensively (40%) polyadenylylated.

To study the mechanism and physiological function of bacterial mRNA polyadenylylation, we have purified a poly(A) polymerase from *E. coli*, identified its gene, and achieved its overexpression by inserting the gene into an expression vector.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. Frozen cells of *E. coli* K-12 grown in rich medium [0.5% yeast extract/4% (wt/vol) NZ amine/0.64% K_2HPO_4/1% glucose] were purchased from Grain Processing (Muscatine, IA). Plasmid pJL89 carrying the *pcnB* locus (17) was the gift of J. S. Parkinson (University of Utah). The expression vector pRE1, its host *E. coli* C600, and *E. coli* MZ (λ *c*I857) (18) were the gift of P. Reddy (National Institute of Standards and Technology).

Enzymes and Chemicals. Most of the restriction endonucleases were from New England BioLabs; *Taq* DNA polymerase was from Perkin–Elmer/Cetus; radioactive nucleotides were from New England Nuclear; rifampicin and phenyl-agarose were from Sigma; Whatman cellulose phosphate P-11, Sephacryl S-200, and poly(A) were from Pharmacia; *E. coli* tRNA was from Schwarz/Mann; and DE-81 discs were from Whatman. Oligonucleotide primers were synthesized by the phosphoramidite method with a MilliGen/BioSearch (Novata, CA) Cyclone DNA synthesizer.

Purification of Poly(A) Polymerase. Frozen *E. coli* cells (100 g) were thawed, suspended in 120 ml of buffer A [50 mM Tris·HCl, pH 7.9/0.2 M KCl/10 mM $MgCl_2$/1 mM EDTA/1 mM dithiothreitol/5% (vol/vol) glycerol/1 mM *p*-toluenesulfonyl fluoride/pancreatic DNase I (1 μg/ml)], and disrupted in a French pressure cell. This and all subsequent steps were carried out at 4°C. The crude extract was centrifuged at 30,000 × *g* for 30 min, and the supernatant was adjusted to 1 M NH_4Cl with solid NH_4Cl and centrifuged for 4 h at 100,000 × *g* to sediment ribosomes. Ammonium sulfate (27.2 g/100 ml) was added to the supernatant solution to give 45% saturation and the resulting precipitate was redissolved in buffer B (25 mM Tris·HCl, pH 7.9/1 mM EDTA/1 mM dithiothreitol/5% glycerol) and dialyzed against buffer B containing 0.3 M NaCl. The solution was diluted with the same buffer to 50 ml prior to applying to a phosphocellulose P-11 column (30 ml, 2 × 10 cm) at a slow rate. After washing with 200 ml of buffer B containing 0.3 M NaCl, the column was developed with a linear gradient (300 ml) from 0.3 M to 1.1 M NaCl in buffer B. Poly(A) polymerase was eluted at 0.7 M NaCl as a major peak, which was pooled and precipitated with ammonium sulfate (28 g/100 ml); a minor peak eluting at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed at: Department of Metabolic Regulation, Boston Biomedical Research Institute, 20 Staniford Street, Boston, MA 02114.

0.5 M NaCl was not included. The concentrated enzyme was subjected to gel filtration on Sephacryl S-200 (85 ml, 1 × 110 cm), equilibrated with buffer S [10 mM Tris·HCl pH 7.9/10 mM $MgCl_2$/1 mM EDTA/1 mM dithiothreitol/500 mM KCl/20% (vol/vol) ethylene glycol]. A single symmetrical peak of poly(A) polymerase activity emerged from the column after ≈58 ml. The active fractions were pooled, dialyzed against 1 M $(NH_4)_2SO_4$ containing 10 mM Tris·HCl (pH 7.9) and 0.2 mM EDTA, and loaded on a phenyl-agarose column (1 ml) equilibrated with the same buffer. The enzyme was eluted by applying a gradient (20 ml) decreasing from 1 M to 50 mM $(NH_4)_2SO_4$ in the same buffer. The enzyme that eluted as a single peak at 0.65 M $(NH_4)_2SO_4$ yielded a double band after SDS/PAGE, the two components being present in equivalent amounts and paralleling the elution of poly(A) polymerase activity from the phenyl-agarose column. Active fractions were brought to 1 mM dithiothreitol and 9.0% glycerol and stored at −10°C.

Protein Sequencing. N-terminal sequences were determined by the methods developed by Matsudaira (19). The poly(A) polymerase peak fractions from phenyl-agarose were combined, dialyzed against 10 mM Tris·HCl, pH 7.4/5 mM 2-mercaptoethanol/1 mM EDTA/0.3 M NaCl, electrophoresed in a gradient polyacrylamide gel (7.5–15%) for complete separation of the doublet bands, which differed in apparent molecular mass by ≈3 kDa, and electroblotted from the gel onto a Problott membrane (Applied Biosystems), and the Coomassie-stained poly(A) polymerase bands were excised. The two bands were sequenced separately by W. Lane (Harvard Microchemistry Facility).

Enzyme Assay. Poly(A) polymerase was assayed in a volume of 50 μl containing 50 mM Tris·HCl (pH 7.9 at 4°C), 10 mM $MgCl_2$, 2 mM $MnCl_2$, 1 mM EDTA, 1 mM dithiothreitol, 0.25 M NaCl, rifampicin (5 mg/ml; to suppress RNA-polymerase activity), 0.2 mM KH_2PO_4 (to suppress polynucleotide phosphorylase activity), 1 mM phospho*enol*pyruvate, 0.3 μg of pyruvate kinase, 25 μg of gelatin, 0.2 mM [^{3}H]ATP (50 μCi/μmol; 1 Ci = 37 GBq), and 12 μg of tRNA or 10 μg of poly(A) as primer (8). Purified enzyme fractions were assayed without rifampicin, KH_2PO_4, and an ATP-regenerating system. After 10 min at 37°C, the reaction was stopped by the addition of 20 mM EDTA and the mixture was spotted on DEAE filter discs. The discs were washed with 0.3 M ammonium formate (pH 8.0) and dried, and radioactivity was measured. One unit is the amount of enzyme that incorporates 1 nmol of AMP per 10 min at 37°C.

DNA Amplification. PCRs were carried out according to Perkin–Elmer/Cetus with 1 μg of template DNA and *Taq* DNA polymerase in a final volume of 100 μl. With highly degenerate primers (300 μg/ml), the reaction after denaturation was started with a low nonstringent annealing temperature of 37°C for 30 min, followed by a slow ramp time of 2.5 min to reach the primer-extension temperature of 72°C, which was maintained for 2 min. This was done for the initial 5 cycles, which were followed by 40 cycles of a 1-min denaturation at 94°C, a 2-min annealing at 37°C with primers, and a 3-min extension at 72°C. Amplification with mismatched primers (400 μg/ml) was run through 40 cycles with the following parameters: 1 min at 94°C, 2 min at 50°C, and 3 min at 72°C. The amplification products were purified by electrophoresis in low-melting-temperature agarose.

Cloning of the *pcnB* Locus in the pRE1 Overexpression Vector. To insert the *pcnB* locus into the unique *Nde* I site of pREI, which is located just downstream of the λ*c*II translation start site, the putative poly(A) polymerase coding sequence was amplified by PCR with mismatched primers (Table 1) designed to generate a *Nde* I site near the initiation codon and a *Bam*HI site after the stop codon. The upstream primer covers the UUG initiation codon with three mismatches, converting it into a *Nde* I site and at the same time into an AUG initiation codon. The downstream primer also has three mismatches just beyond the translation stop codon, generating a *Bam*HI site. Amplification was carried out with 1 μg of pJL89 DNA and the amplified fragment was cloned into the *Nde* I and *Bam*HI sites of the expression vector pREI (20), to yield the *pcnB* recombinant plasmid pRE1-1. Recombinant plasmids with truncated *pcnB* inserts (see Fig. 2) were constructed in an analogous manner by using the primers listed in Table 1. The insert of plasmid pRE1-2 starts at the first AUG codon in the *pcnB* open reading frame, 174 residues downstream of the putative UUG start codon (see Fig. 1). The insert of pRE1-3 is truncated at the 3′ end by the generation of a UAG codon at residue 1202, and the insert of pRE1-4 combines the truncations of plasmids pRE1-2 and pRE1-3, whereas the insert of pRE1-5 has the full upstream coding sequence but is truncated even further downstream by the introduction of a UAG codon at residue 1034. The vector pRE1 and its various *pcnB* recombinants were used to transform *E. coli* MZ1. The resulting transformants were grown at 30°C and λP_L expression was induced by temperature shift to 39°C. After 1 hr at 39°C, cells were harvested, disrupted with a French pressure cell, and assayed for poly(A) polymerase activity either directly or after centrifugation at 100,000 × *g*. Samples of the cell extracts were also subjected to PAGE (10% gel) in the presence of SDS, followed by Coomassie blue staining.

DNA Sequencing. The nucleotide sequence of the region upstream of *pcnB* in plasmid pJL89 was determined by the dideoxynucleotide chain-termination method using Sequenase Version 2.0 (United States Biochemical) and specific primers (5′-GTTTTCCCAGTCACGAC and 5′-TCCGGCTCGTATGTTGTG). The double-stranded DNA template was first denatured with alkali.

RESULTS

Purification and Properties of Poly(A) Polymerase. The partial purification of poly(A) polymerase from extracts of *E. coli* had been reported by several laboratories (6–9). Our purification procedure was in part based on these earlier studies but employed additional steps such as gel filtration on Sephacryl S-200 and hydrophobic chromatography on phenyl-Sepharose and resulted in an at least 90% pure enzyme preparation, purified 920-fold over the 100,000 × *g* supernatant with a 3.4% yield (Table 2). The properties of the enzyme were studied in partially purified preparations. Poly(A) polymerase activity was completely dependent on the presence of an RNA primer, tRNA being ≈10 times more effective than

Table 1. Primers for the amplification of *pcnB* segments for cloning in pRE1

Plasmid	Upstream primer (*Nde* I site)	Downstream primer (*Bam*HI site)
pRE1-1	5′-GTCCTGAATGA**CATATG**ACACTACCGAGGTG	5′-CTGCCTATGGCA**GG**AT**C**CGCCACTGTCATG
pRE1-2	5′-ATGCCCTGAAG**CATATG**TACAGGCTCAATAA	5′-CTGCCTATGGCA**GG**AT**C**CGCCACTGTCATG
pRE1-3	5′-GTCCTGAATGA**CATATG**ACACTACCGAGGTG	5′-AGCAG**GA**TCCATGCGCGTTTACCCTGACGACGGGAC**TA**ACGC
pRE1-4	5′-ATGCCCTGAAG**CATATG**TACAGGCTCAATAA	5′-AGCAG**GA**TCCATGCGCGTTTACCCTGACGACGGGAC**TA**ACGC
pRE1-5	5′-GTCCTGAATGA**CATATG**ACACTACCGAGGTG	5′-CTGTGCCG**GA**TCCAGCAGTGG**C**TACCAGAACA

Underlined residues differ from the genomic DNA sequence and were altered to generate appropriate restriction endonuclease cleavage sites and/or translation start or termination sites (boldface type).

Table 2. Purification of poly(A) polymerase from *E. coli*

Fraction	Protein, mg	Enzyme units, nmol/10 min	Specific activity, units/mg	Purification, fold	Yield, %
Supernatant (100,000 × *g*)	5160	48,000	9.4	1	100
Ammonium sulfate	1680	38,000	23	2.4	78
Phosphocellulose	39	6,500	164	18	13
Sephacryl S-200	0.8	3,200	4010	427	6.6
Phenyl-agarose	0.19	1,640	8630	918	3.4

Poly(A) polymerase was purified from 100 g of cells.

poly(A), and also on the presence of divalent cations, with optimum activities at 10 mM Mg^{2+} and 2 mM Mn^{2+} (data not shown). The enzyme aggregated at low ionic strength and its activity was stimulated by salt, with optimum activity at 300 mM NaCl but inhibition at higher concentrations (data not shown). The enzyme was not inhibited by rifampicin or streptolydigin (gift of G. B. Whitfield, Upjohn) at concentrations that completely inhibit RNA polymerase but was inhibited completely by 10 mM aurintricarboxylate (data not shown).

N-Terminal Sequence of Poly(A) Polymerase. The most highly purified preparation of poly(A) polymerase from *E. coli* K-12 yielded a doublet band after SDS/PAGE. The difference in electrophoretic mobility of the two components, which were present in equivalent amounts and paralleled the elution of poly(A) polymerase activity from the phenyl-Sepharose column, corresponded to an apparent molecular mass difference of 3 kDa, and may have been caused by limited C-terminal proteolysis or some other type of covalent modification. The average molecular mass of the bands was ≈55 kDa, and N-terminal analysis of each component of the doublet yielded the same 25-residue sequences, with the first 15 residues at a high confidence level: H_2N-Lys-Val-Leu-Ser-Arg-Glu-Glu-Ser-Glu-Ala-Glu-Gln-Ala-Val-Ala-Arg-Arg-Pro-Gln-Val-Val-Ile-Pro-Arg-Glu.

Amplification of *E. coli* Genomic DNA with Degenerate Oligonucleotide Probes. The determination of the sequence of 25 N-terminal amino acid residues of *E. coli* poly(A) polymerase provided the necessary information for the synthesis of degenerate oligonucleotide probes. Initial attempts to use such probes for screening an *E. coli* genomic EMBL4 library (21) proved impractical because multiple positive clones were found that had only partial homology to the sequence of interest. We therefore turned to an alternative approach that involved the use of two sets of degenerate probes, corresponding to 6 or 7 amino acid residues, at the two extremes of the 25-residue sequence, to amplify the corresponding sequence from *E. coli* genomic DNA. The primers corresponding to the N-terminal sequence were a mixture of 5′-GGTCTAGAA(GA)GT(GCT)(TC)TGAG(TC)CG(TC)-GA(GA) (primer 1) and GGTCTAGAA(GA)GT(GCT)(TC)T-GTC(CTG)CG(TC)GA(GA)GA (primer 2) and contained *Xba* I sites (underlined). The primer corresponding to the C-terminal sequence was 5′-CCGAATTC(GA)CG(CTG)GG-(GA)AT(CTG)AC(CTG)AC(CT)TG (primer 3) and contained an *Eco*RI site (underlined). The combination of primers 1 and 3 led to the amplification of *E. coli* DNA, whereas the combination of primers 2 and 3 was ineffective for amplification. The PCR product was digested with *Eco*RI and *Xba* I, cloned into pUC18, and sequenced. The nucleotide sequence of a clone with the 74-base-pair insert, differed from the empirically determined 25-residue N-terminal sequence by the absence of an arginine at residue 34 and the presence of a threonine at residue 37 (Fig. 1).

Identification of the *E. coli pcnB* Locus as the Coding Region for Poly(A) Polymerase. A search of the GenBank data base (June 1991) for sequences corresponding to the PCR product revealed an exact match in the upstream portion of the *E. coli pcnB* locus (17). The alignment of the putative poly(A) polymerase coding sequence with the *E. coli pcnB* gene is shown in Fig. 1. The N-terminal poly(A) polymerase coding sequence matches a portion of the *pcnB* open reading frame upstream from the translation start site postulated by Liu and Parkinson (17), suggesting that the translation start site may actually be further upstream, perhaps at the UUG codon 17 residues upstream from the observed N terminus of the purified enzyme. The 0.25-kilobase nucleotide sequence upstream of the putative translation start site was determined but does not contain any remarkable elements such as extensive open reading frames, regions of dyad symmetry, or obvious transcription control elements (Fig. 1).

To determine whether *pcnB* indeed encodes a poly(A) polymerase, we examined the levels of this enzyme in bacterial strains carrying multiple copies of this gene. How-

```
   -211       -201       -191       -181       -171       -161       -151
AGGCGC TAAACTTTCC AAGCAGAATC ATGCGCCTGC GTTGCCGAAA GGCGATCCAC GCCCGGTACT

   -141       -131       -121       -111       -101        -91        -81
AATCGCGGCA CTTCAATTTC TGGGGCAGCA GGCAGAAGCA CACTGGCAGG ATTTCAGCGT CGAGCAAATC

    -71        -61        -51        -41        -31        -21        -11
CTTCAGTCAG CCGTCAAAAA CTGGGCTTAC CGCCGTGCCC GAGTCGGCAA TTGTAAATTC AACATTCTCA

     -1         1         10         20         30         43
AATGCGTCAT      GCTGAGCTAT GATTAGCCGC TATTTTTTTG TCCTGAATGATGT
                                                 ****
                                                                        94
   TTG ACA CTA CCG AGG TGT ACT ATT TTT ACC CGA GTC GCT AAT TTT TGC CGC
  Fmet Thr Leu Pro Arg Cys Thr Ile Phe Thr Arg Val Ala Asn Phe Cys Arg
     1                                  10
                                                                       148
AAG GTG CTA AGC CGC GAG GAA AGC GAG GCT GAA CAG GCA GTC GCC CGT CCA CAG
Lys Val Leu Ser Arg Glu Glu Ser Glu Ala Glu Gln Ala Val Ala Arg Pro Gln
       20                                  30
                                                                       202
GTG ACG GTG ATC CCG CGT GAG CAG CAT GCT ATT TCC CGC AAA GAT ATC AGT GAA
Val Thr Val Ile Pro Arg Glu Gln His Ala Ile Ser Arg Lys Asp Ile Ser Glu
               40                                              50
                                                                       256
AAT GCC CTG AAG GTA ATG TAC AGG CTC AAT AAA GCG GGA TAC GAA GCC TGG CTG ...
Asn Ala Leu Lys Val Met Tyr Arg Leu Asn Lys Ala Gly Tyr Glu Ala Trp Leu ...
                       60                                          70
```

FIG. 1. Nucleotide sequence of the upstream region of the *E. coli pcnB* locus, with the its deduced translation product. The underlined nucleotides correspond to the experimentally determined N-terminal sequence of poly(A) polymerase. The double-underlined methionine residue is the translation start site proposed by Liu and Parkinson (17). The asterisks indicate a potential ribosome binding site. The numbering system used is that of Liu and Parkinson (17), with nucleotides upstream of those sequenced earlier (17) indicated by negative numbers. For residues 1–256, in which the sequence determined here overlaps with that determined earlier (17), the two sequences are in complete agreement.

ever, it was not possible to clone *pcnB* into conventional inducible expression vectors (e.g., pTZ-7, constructed and kindly provided by Stanley Tabor, Harvard Medical School), suggesting that even the uninduced basal level of expression of the enzyme may be highly toxic to the host. Accordingly, we selected a vector (pREI) designed specifically for the expression of lethal gene products (20). pRE vectors are based on the λP_L promoter and λcII ribosome binding site including an ATG translation initiation codon, and two transcription terminators (t_o and t_{RI}) positioned 5′ and 3′ to the λP_L promoter, respectively, to eliminate nonspecific transcription and reduce leaky P_L transcription. Induction of the λP_L promoter in an *E. coli* N^+ background antiterminates transcription at t_{RI} allowing full transcription of downstream genes. The coding sequence of the *E. coli pcnB* locus was cloned into pRE1 to form the recombinant plasmid pRE1-1. Plasmid expression was induced in strains of *E. coli* MZ1 with pRE1 or pRE1-1, and extracts were assayed for poly(A) polymerase activity by using either tRNA or poly(A) as primer. Enzyme activity was ≈150-fold higher in the 100,000 × *g* supernatant fraction from the strain harboring the *pcnB* recombinant plasmid than from the control strain. The overexpressed enzyme had a specific activity of 1100 units per mg of protein in the unpurified 100,000 × *g* supernatant fraction, compared to 8600 units per mg of protein for the homogenous poly(A) polymerase preparation from the wild-type strain, and its activity was 8-fold greater with tRNA than with poly(A) as primer.

To correlate the capacity to overexpress poly(A) polymerase activity with the open reading frame constituting the *pcnB* gene, pRE1 derivatives were constructed with *pcnB* inserts that were truncated at either end of the open reading frame, as illustrated in Fig. 2. After induction of the modified *pcnB* genes, cell samples were disrupted in a French pressure cell and the extracts were subjected to denaturing PAGE and assayed for poly(A) polymerase activity. The PAGE patterns obtained with extracts of cells transformed with recombinant plasmids showed additional protein components that were absent from control cells transformed with nonrecombinant pRE1. The molecular masses of the additional components roughly corresponded to those of the predicted truncated PcnB proteins (Fig. 2). Although cells transformed with any of the modified constructs produced a PcnB-specific polypeptide product, only extracts from cells transformed with the full-length *pcnB* coding region had greatly elevated (380-fold in uncentrifuged extracts) poly(A) polymerase activity. A truncated protein lacking 58 C-terminal amino acid residues retained some residual activity, whereas deletion of an additional 168 base pairs from the C-terminal end or deletion of the 174 base pairs from the N-terminal end led to complete loss of poly(A) polymerase expression (Fig. 2*B*). The N-terminal deletion is of special interest because its N terminus corresponds to the first in-phase AUG codon that had originally been postulated by Liu and Parkinson (17) to be the N terminus of the PcnB protein.

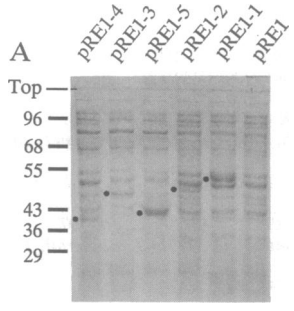
DISCUSSION

We report the purification to near homogeneity of a poly(A) polymerase from *E. coli*. The ≈1,000-fold purified enzyme preparation migrates as a doublet on denaturing PAGE, but both components of the doublet have the same N-terminal amino acid sequence, indicating that they are closely related proteins. By using the PCR with degenerate oligonucleotide primers corresponding to hexapeptide sequences flanking the two extremes of the tentative 25-residue N-terminal sequence of poly(A) polymerase, it was possible to identify a 74-base-pair genomic fragment encoding this portion of the protein. This sequence exactly matched a region associated with the *pcnB* locus, a gene involved in the regulation of plasmid copy number, which was originally identified by Lopilato *et al.* (22) and subsequently cloned and sequenced by Liu and Parkinson (17). Interestingly, the sequence identified as corresponding to the N terminus of the purified poly(A) polymerase lies upstream of the first AUG codon in the open reading frame defining the *pcnB* locus, proposed by Liu and Parkinson (17) to be the translation start site of the PcnB protein. Instead a UUG codon at position 44 in the extreme upstream portion of the *pcnB* locus is the most plausible candidate (Fig. 1).

Although UUG is a relatively rare translation start codon, it is used for certain RNA processing genes in *E. coli*, such as those for polynucleotide phosphorylase (23) and RNase D.† UUG initiator codons are also found in the genes for adenylate cyclase (24), DeaD (25), ribosomal protein S20 (26), and NADH dehydrogenase (27) and perhaps constitute a mechanism for limiting expression. There is no strong Shine–Dalgarno (28) ribosome binding site upstream of the UUG codon in the *pcnB* gene, but the sequence TGAT at position 38 is comparable to the ribosome binding site in the *deaD* gene (25).

If translation of poly(A) polymerase indeed started at the UUG codon at position 44, the observed N terminus of the purified enzyme would have to be the result of proteolytic processing, perhaps by cleavage between an arginine and a lysine. It is interesting that the outer membrane-associated OmpT protease preferentially cleaves arginyl–lysine bonds

†Zhang, J. & Deutscher, M. P., ASM Conference on RNA Processing and mRNA Decay in Procaryotic Cells, Oct. 27–29, 1991, North Falmouth, MA.



B

Plasmid	Plasmid insert and predicted translation product	Molecular weight of translation product: Predicted (processed)	Molecular weight of translation product: Observed	Poly(A) polymerase activity (units/mg protein)
pRE1-1	44 TTG, 218 ATG, 1034 TAC→TAG, 1202 ATG→TAG, 1458 TAG	53,871 (51,859)	53,000	230
pRE1-2		47,263	48,000	0.8
pRE1-3		44,330 (42,318)	46,000	3.7
pRE1-4		37,224	38,000	0.4
pRE1-5		38,122 (36,110)	39,000	0.3
pRE1	No insert	None	None	0.6

Processing site

FIG. 2. Overproduction of the products of the normal and truncated versions of the *pcnB* locus in *E. coli* MZ1 transformed with recombinant pRE1 plasmids. Cells were disrupted in a French pressure cell and the crude extracts were analyzed by electrophoresis and enzyme assay. (*A*) SDS/PAGE. Equal amounts of protein (12 μg) were subjected to electrophoresis in 10% polyacrylamide gels with standards of known molecular mass (shown in kDa to the left), followed by staining with Coomassie blue. Protein bands absent from the pRE1- transformed control culture are indicated by dots to the left of the lanes. (*B*) Diagram of the plasmid inserts, predicted molecular masses of the translation products of the cloned *pcnB* fragments, with the products processed by cleavage before Lys-18 in parentheses, observed molecular weights of the *pcnB*-encoded translation products, and measured poly(A) polymerase activity. Nucleotides are numbered according to the system of Liu and Parkinson (17).

(29). This protease is insensitive to the protease inhibitor phenylmethylsulfonyl fluoride (29), which was used as a protective agent in our purification procedure, and it is thus possible that the 17 N-terminal amino acid residues of poly(A) polymerase were lost during the early stages of purification. Whether the N terminus of the purified poly(A) polymerase is an artifact of purification or whether it is generated by *in vivo* proteolysis remains to be determined.

What is the evidence that the *pcnB* locus indeed encodes poly(A) polymerase? Transformation of *E. coli* with the expression vector pRE1 carrying the entire *pcnB* coding sequence leads, upon induction, to the overproduction of a 53-kDa polypeptide and 300-fold elevated levels of poly(A) polymerase (Fig. 2). Deletion of 10–20% of the nucleotides from either end of the *pcnB* open reading frame leads to the overexpression of correspondingly truncated polypeptide chains, which are devoid of poly(A) polymerase activity. These observations support the conclusion that the 53-kDa polypeptide encoded by the *pcnB* gene corresponds to poly(A) polymerase. Moreover, the fact that deletion of nucleotides 44–217 in plasmid pRE1-2 completely abolished expression of poly(A) polymerase activity (Fig. 2) suggests that the translation start site for poly(A) polymerase lies at the very end of the gene, presumably at the UUG codon (nucleotides 44–46) and not at the first in-frame AUG codon (nucleotides 218–220), which had originally been postulated as the translation initiation site of the *pcnB* gene product (17). A very low but significant level of poly(A) polymerase activity was observed in *E. coli* transformed with plasmid pRE1-3, in which 0.25 kilobase had been deleted from the downstream end of the *pcnB* gene, suggesting that the C-terminal portion of poly(A) polymerase may not be absolutely essential for activity.

Since poly(A) polymerase is encoded by the *pcnB* locus, it is of interest to summarize what is known about this gene: *pcnB* was originally isolated as a low-plasmid-copy-number host mutation, mapping at 3 min on the *E. coli* chromosome (22). It was subsequently studied with *cheW* as an indicator by Liu and Parkinson (17, 30), who found that *pcnB* mutations are recessive and that the *pcnB* locus cannot be cloned in multicopy plasmids, suggesting that large amounts of its product are lethal. Disruption of the *pcnB* locus results in a 67% reduction of growth rates in rich medium (17). It seems, therefore, that *pcnB* encodes a trans-acting product whose expression is set at a relatively low level.

Whether the poly(A) polymerase encoded by *pcnB* is directly involved in the control of plasmid copy number is a question that can only be addressed after the molecular mechanisms that control plasmid replication are more fully understood. It has been suggested that the *pcnB* locus encodes a protein with multiple functions (17). This prediction could be tested by comparing the levels poly(A) polymerase activity and plasmid copy number control in deletion mutants of the type described in Fig. 2. In this connection, it should be noted that the shorter of our C-terminal deletions (plasmid pRE1-3) that can express low but significant poly(A) polymerase activity corresponds to deletion pJL89Δ3 and mini-kam insertion pJL89k2, constructed by Liu and Parkinson (17), which have a normal plasmid copy number control. On the other hand, the longer of our C-terminal deletions (plasmid pRE1-5) that completely abolishes poly(A) polymerase expression corresponds to Liu and Parkinson's constructs pJL89Δ4 and pJL89inv1 (17), which have lost the plasmid copy number control function.

We have no direct evidence that mRNA polyadenylylation in *E. coli* is carried out by the poly(A) polymerase we have purified. However, preliminary measurements of the level of polyadenylylation of mRNA in permeable cells have shown a 50% reduction in a mutant with a deletion in the *pcnB* locus (ref. 30 and unpublished data). Thus, the poly(A) polymerase purified by us may contribute significantly to mRNA polyadenylylation in *E. coli* but is not the only enzyme that does so. Sequence comparison of the deduced PcnB product with other proteins reveals a region with high similarity to a segment of *E. coli* tRNA nucleotidyltransferase (32) that may be an RNA binding site. It is interesting that the properties of our purified *E. coli* poly(A) polymerase and of others (6–9) revealed striking similarities to those reported for the mammalian enzyme involved in the polyadenylylation of mRNA (2–5, 31): Both enzymes are monomeric, with molecular masses of ≈55 kDa; both enzymes have a strong tendency to aggregate under certain ionic conditions; both enzymes have a pH optimum near pH 8 and require Mg^{2+} or Mn^{2+} for activity; both enzymes are relatively unspecific with respect to the primer for poly(A) synthesis; both enzymes consist of isoforms as indicated by the appearance of a doublet band upon SDS/PAGE; and both enzymes are subject to partial proteolysis during purification. On the other hand, in spite of the many similarities in the properties of the poly(A) polymerases from different sources, there is no obvious sequence similarity between the *E. coli* poly(A) polymerase and the corresponding enzymes from calf thymus (2), vaccinia virus (4), and yeast (5). It remains to be determined whether prokaryotic and eukaryotic poly(A) polymerases have analogous physiological functions.

We thank Drs. J. S. Parkinson, P. Reddy, and S. Tabor for the generous gifts of plasmids and bacterial strains and Dr. Henry Paulus for the critical reading of the manuscript. This work was supported by Grant GM-26517 from the National Institute of General Medical Sciences.

1. Wickens, M. (1990) *Trends Biochem. Sci.* **15**, 277–281.
2. Raabe, T., Bollum, F. J. & Manley, J. L. (1991) *Nature (London)* **353**, 229–234.
3. Wahle, E., Martin, G., Schiltz, E. & Keller, W. (1991) *EMBO J.* **10**, 4251–4257.
4. Gershon, P. D., Ahn, B.-Y., Garfield, M. & Moss, B. (1991) *Cell* **66**, 1269–1278.
5. Lingner, J., Kellerman, J. & Keller, W. (1991) *Nature (London)* **354**, 496–498.
6. August, J. T., Ortiz, P. J. & Hurwitz, J. (1962) *J. Biol. Chem.* **237**, 3786–3793.
7. Schafer, R., Zillig, W. & Priess, H. (1972) *FEBS Lett.* **25**, 87–90.
8. Sippel, A. E. (1973) *Eur. J. Biochem.* **37**, 31–40.
9. Ramanarayanan, M. & Srinivasan, P. R. (1976) *J. Biol. Chem.* **251**, 6274–6286.
10. Sarkar, N., Langley, D. & Paulus, H. (1978) *Biochemistry* **17**, 3468–3474.
11. Gopalakrishna, Y. & Sarkar, N. (1982) *Biochemistry* **21**, 2724–2729.
12. Gopalakrishna, Y., Langley, D. & Sarkar, N. (1981) *Nucleic Acids Res.* **9**, 3545–3554.
13. Gopalakrishna, Y. & Sarkar, N. (1982) *J. Biol. Chem.* **257**, 2747–2750.
14. Karnik, P., Gopalakrishna, Y. & Sarkar, N. (1986) *Gene* **49**, 161–165.
15. Taljanidisz, J., Karnik, P. & Sarkar, N. (1987) *J. Mol. Biol.* **193**, 507–515.
16. Karnik, P., Taljanidisz, J., Sasvari-Szekely, M. & Sarkar, N. (1987) *J. Mol. Biol.* **196**, 347–354.
17. Liu, J. & Parkinson, J. S. (1989) *J. Bacteriol.* **171**, 1254–1261.
18. Zuber, M., Patterson, T. A. & Court, D. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4514–4518.
19. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038.
20. Reddy, P., Peterkofsky, A. & McKenney, K. (1989) *Nucleic Acids Res.* **17**, 10473–10488.
21. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.
22. Lopilato, J., Bortner, S. & Beckwith, J. (1986) *Mol. Gen. Genet.* **205**, 285–290.
23. Regnier, P., Grunbery-Manago, M. & Portier, C. (1987) *J. Biol. Chem.* **262**, 63–68.
24. Roy, A., Haziza, C. & Danchin, A. (1983) *EMBO J.* **2**, 791–797.
25. Toone, W. M., Rudd, K. E. & Friesen, J. D. (1991) *J. Bacteriol.* **173**, 3291–3302.
26. Mackie, G. A. (1981) *J. Biol. Chem.* **256**, 8177–8182.
27. Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A. & Shaw, D. C. (1981) *Eur. J. Biochem.* **116**, 165–170.
28. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
29. Sugimura, K. & Higashi, N. (1988) *J. Bacteriol.* **170**, 3650–3654.
30. Liu, J. & Parkinson, J. S. (1991) *J. Bacteriol.* **173**, 4941–4951.
31. Wahle, E. (1991) *J. Biol. Chem.* **266**, 3131–3139.
32. Masters, M., March, J. B., Oliver, I. R. & Collins, J. F. (1990) *Mol. Gen. Genet.* **220**, 341–344.