Characterization of a Novel DNA Primase from the *Salmonella typhimurium* Bacteriophage SP6†

Timothy Y. Tseng, David N. Frick, and Charles C. Richardson*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received September 17, 1999; Revised Manuscript Received November 9, 1999

ABSTRACT: The gene for the DNA primase encoded by *Salmonella typhimurium* bacteriophage SP6 has been cloned and expressed in *Escherichia coli* and its 74-kDa protein product purified to homogeneity. The SP6 primase is a DNA-dependent RNA polymerase that synthesizes short oligoribonucleotides containing each of the four canonical ribonucleotides. GTP and CTP are both required for the initiation of oligoribonucleotide synthesis. In reactions containing only GTP and CTP, SP6 primase incorporates GTP at the 5′-end of oligoribonucleotides and CMP at the second position. On synthetic DNA templates, pppGpC dinucleotides are synthesized most rapidly in the presence of the sequence 5′-GCA-3′. This trinucleotide sequence, containing a cryptic dA at the 3′-end, differs from other known bacterial and phage primase recognition sites. SP6 primase shares some properties with the well-characterized *E. coli* bacteriophage T7 primase. The T7 DNA polymerase can use oligoribonucleotides synthesized by SP6 primase as primers for DNA synthesis. However, oligoribonucleotide synthesis by SP6 primase is not stimulated by either the *E. coli*- or the T7-encoded ssDNA binding protein. An amino acid sequence alignment of the SP6 and T7 primases, which share only 22.4% amino acid identity, indicates amino acids likely critical for oligoribonucleotide synthesis as well as a putative Cys_3His zinc finger motif that may be involved in DNA binding.

DNA polymerases cannot initiate the de novo synthesis of both new strands (*1*). All known DNA polymerases require a priming mechanism, the most common being the use of short oligoribonucleotides annealed to ssDNA.[1] DNA polymerases can efficiently add deoxyribonucleotides to the 3′-ends of oligoribonucleotides that are themselves synthesized in a 5′ to 3′ template-dependent manner by a class of enzymes called DNA primases (*2*). Whereas other types of polymerases will read any sequence as a template, primases are unique in that they synthesize oligoribonucleotides primarily at DNA sequences designated as primase recognition sites. The well-characterized bacterial and phage primases each recognize a unique trinucleotide DNA sequence. For example, bacteriophage T7 primase synthesizes oligoribonucleotides at primase recognition sites that begin with the sequence 5′-GTC-3′ (*3, 4*). Oligonucleotide synthesis begins opposite the dT in the recognition site, and all primers start with the dinucleotide pppApC (*5, 6*). The 3′-dC in the recognition site is thus designated as "cryptic" because although essential, its complement is not part of the primer (*3, 7*). Other primases synthesize oligoribonucleotides at different trinucleotide sequences also containing cryptic nucleotides at the 3′-ends. For example, the *Escherichia coli* DnaG primase recognizes the trinucleotide 5′-CTG-3′ (*8*), and the T4 gp61 primase recognizes 5′-G(T/C)T-3′ (*9*). A third class of enzymes called DNA helicases is equally essential for the replication of duplex DNA. Helicases produce the ssDNA templates required by primases and polymerases by unwinding the duplex DNA at the replication fork. Generally, replicative DNA helicases are hexameric proteins that separate the double helix through unidirectional translocation along DNA using energy derived from the hydrolysis of NTPs. The *E. coli* DnaB protein (*10, 11*) is the prototype of this class of helicases.

In the cell, polymerase, primase, and helicase proteins must work together to rapidly synthesize both strands of DNA. The resulting complex macromolecular machine is often referred to as the replisome. The replisome is composed of nucleic acids and multiple proteins. In *E. coli*, for example, the replisome is composed of over 20 different proteins (*1*). To start DNA synthesis on either strand, the primase and polymerase must function together. On the leading strand, DNA is synthesized in a continuous manner requiring only a single instance of priming. The replication of the antiparallel lagging strand, on the other hand, must frequently be reinitiated as the DNA is progressively unwound by the helicase. The resulting discontinuous synthesis gives rise to small fragments of DNA (Okazaki fragments). Primases are also associated functionally and physically with DNA helicases. In most organisms, primase and helicase functions are encoded by separate genes and translated into distinct proteins. When primase activity is measured in vitro, the most efficient oligoribonucleotide synthesis occurs only when the

† This work was supported by United States Public Health Service Grant AI-06045 and by American Cancer Society Grant NP-1Z to C.C.R. D.N.F. was supported by American Cancer Society Postdoctoral Fellowship PF-4452.

* To whom correspondence should be addressed at the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115. Tel.: 617-432-3130; Fax: 617-432-3362; E-mail: ccr@hms.harvard.edu.

[1] Abbreviations: DTT, dithiothreitol; (d)NTP(s), (deoxy)nucleoside triphosphate(s); (d)NMP(s), (deoxy)nucleoside monophosphate(s); IPTG, isopropyl-β-D-thiogalactopyranoside; nt, nucleotide(s); LB, Luria Bertani medium; PCR, polymerase chain reaction; ssDNA, single-stranded DNA.

10.1021/bi992155t CCC: $19.00
Published on Web 01/27/2000

helicase is added. This stimulation results from the physical interaction of the two proteins, and such an interaction has been demonstrated for the *E. coli* DnaG primase and the *E. coli* DnaB helicase (*12, 13*). Likewise in phage T4, the gp41 protein (helicase) combines in a 6:1 complex with the T4 gp61 protein (primase) when bound to DNA and ATP (*14*−*16*). The physical coupling of primase to helicase allows the primase to use the unidirectional translocation of the helicase to locate primase recognition sites in DNA (*17*).

The bacteriophages have long provided attractive models for studying DNA replication because of the relative ease with which one can attain large amounts of phage enzymes. The replisome of phage T7, for example, can be reconstructed in vitro from only four proteins: the T7 gene 5 DNA polymerase; its processivity factor *E. coli* thioredoxin; the T7 gene 2.5 ssDNA binding protein; and the T7 gene 4 primase/helicase (*18*). The T7 phage has made efficient use of its limited genome by combining the helicase and primase functions into the product of a single gene. This one gene, T7 gene 4, encodes two collinear proteins—a 63-kDa form possessing both primase and helicase activity, and a 56-kDa form possessing only helicase function—that differ only in their translational start sites (*19*−*21*). An in-frame translational start site downstream from the start site of the large form encodes the shorter form (*19*). The N-terminal 63 amino acid residues not present in the 56-kDa protein contain a Cys_4 zinc finger essential for primer synthesis (*22*−*24*). Since the 63-kDa gene 4 protein contains both helicase and primase activities, the 56-kDa gene 4 protein is not necessary for phage growth (*25*). It has been proposed that the T7 gene 4 evolved following a fusion of two ancestral genes that were the common ancestors of the current bacterial and bacteriophage primase and helicase genes (*26*).

Only two other proteins are known that combine primase and helicase activities into a single polypeptide; these are encoded by bacteriophages T3 and P4. T3 is a very close relative to T7, and expression of the T3 helicase/primase from a plasmid vector in *E. coli* allows the growth of T7 phage lacking gene 4.[2] Moreover, a chimeric T3/T7 gene 4 protein is a functional primase in vivo and synthesizes oligoribonucleotide at the same recognition sequence as the T7 primase (*23*). This genetic and biochemical evidence indicates that the T3 gene 4 protein can substitute for its T7 homologue, a result that is not surprising considering the extremely high homology (80.8% identity) between the T7 and T3 gene 4 proteins. The 84-kDa α protein of P4, which is 18.8% identical to T7 gene 4 protein, possesses primase and helicase functions and can also recognize the P4 origin of replication (*27*).

Preliminary sequence analysis of the genome of the *Salmonella typhimurium* bacteriophage SP6[3] has uncovered a fourth potential primase/helicase protein (GenBank accession no. AF159357) (*28*). A 74-kDa SP6 gene product, which we have designated gp74, possesses an amino acid sequence that is 22.4% identical to that of the T7 gene 4 protein. The SP6 phage, a morphologically similar, yet distant relative of the T7 phage, has been used extensively to study RNA transcription (*29*−*31*). The putative 74-kDa SP6 primase/helicase contains the six DnaG-like primase and five DnaB-like helicase amino acid signature sequences that are widely conserved among such proteins (*26*). However, *E. coli* strains carrying the SP6 gene do not permit the growth of T7 phage lacking gene 4.[2] These observations raise the intriguing possibility that SP6 gp74 may be both a helicase and a primase protein, and yet, because it is distantly related to the T7 gene 4 protein, it cannot functionally substitute for the T7 gene 4 protein. To examine the biochemical similarities between these two proteins, we have cloned the SP6 primase gene, overexpressed SP6 gp74 in *E. coli*, and characterized the purified SP6 protein. An analysis of substrate and template requirements reveals the SP6 gp74 is a unique DNA primase and not simply a homologue of the T7 gene 4 protein.

EXPERIMENTAL PROCEDURES

DNA, Nucleotides, Enzymes, and Strains. Bacteriophage M13mp18 and ΦX174 ssDNA were obtained from New England Biolabs. Synthetic oligonucleotides were purchased from Integrated DNA Technologies and the Biopolymers Laboratory of Harvard Medical School. T7 gene 2.5 protein was obtained from James M. Stattel (Harvard Medical School), and *E. coli* ssDNA binding protein was purchased from Amersham Pharmacia Biotech. Stanley Tabor (Harvard Medical School) provided purified T7 DNA polymerase and the 63-kDa T7 gene 4 protein. In the 63-kDa form of the T7 gene 4 protein used in these experiments, glycine was substituted for methionine-64 to prevent the translation of the 56-kDa protein. The enzymatic activities of the purified 63-kDa M64G gene 4 protein are indistinguishable from those of the wild-type 63-kDa protein (*25*). T4 DNA ligase, T4 polynucleotide kinase, and [^{32}P]NTPs were purchased from Amersham Pharmacia Biotech. *E. coli* NovaBlue and HMS174(DE3) cells were obtained from Novagen.

Cloning of the SP6 gp74 Gene. Two oligonucleotide primers, SP6NCO (5′-CGCGC GCCAT GGCTA TTATT AACAA TATTC CGTGC CCTGC C-3′) and BAMHI-2 (5′-GCGCG CGGAT CCTCA TCCAT TAAAC TCCTG TGTGT TCTCT T-3′), were used to amplify the gene encoding SP6 gp74 from phage SP6 DNA in reactions containing Deep Vent Polymerase (New England Biolabs) and to attach *Nco*I and *Bam*HI restriction sites (underlined) to the 5′- and 3′-ends of the PCR gene products. The 2010 bp product was purified from an agarose gel using the GeneClean kit (Bio 101), digested with *Nco*I and *Bam*HI, purified, and ligated into the *Nco*I and *Bam*HI sites of a similarly prepared pET24d vector plasmid (Novagen) to create plasmid pET-gp74. The ligation reactions were used to transform NovaBlue competent cells (Novagen). Transformants selected on LB−kanamycin plates were grown in 50 mL of LB medium containing 100 μM kanamycin. Plasmid DNA was recovered and used to transform *E. coli* HMS174(DE3) cells. Induction of HMS174(DE3) cells carrying the plasmid pETgp74 with isopropyl-β-D-thiogalactopyranoside (IPTG) resulted in the expression of a 74-kDa protein. Plasmid pETgp74 was purified from cells after induction using the Rapid Pure Midiprep Automated Fluorescence Sequencing kit (Bio 101). To confirm that no mutations were present in the expressed gene, both DNA strands of pETgp74 were sequenced using the ThermoSequenase cycle sequencing kit (Amersham) with the following primers: 5′-CTACT CGGAC GTCAT AAA-3′, 5′-CTAGA TGTCC AAACT TGA-3′, 5′-GACTT GC-

[2] E. C. Korngold and C. C. Richardson, unpublished data.
[3] J. Rush and C. C. Richardson, unpublished data.

CGG GGAAC AGGC-3′, 5′-TTCAC TACTT CGCGG AGG-3′, 5′-CAATG ATGAT GTTAG AAATA CCC-3′, 5′-TAATA CGACT CACTA TAGGG-3′, 5′-GCTAG TTATT GCTCA GCGG-3′, 5′-CGCTT CGAAG GTGGC AAACT TG-3′, 5′-GTTCA AGAAG ATTAT GTGGG GC-3′, 5′-CCGTA AGCAT CAGCT AATCA TCG-3′, 5′-CTATT CTATG GAGAA GGTAG-3′.

Purification of SP6 gp74. A single colony of *E. coli* strain HMS174(DE3) containing plasmid pETgp74 was inoculated into 50 mL of LB medium containing 100 μM kanamycin. After overnight incubation at 37 °C, these cells were transferred to 4 L of fresh LB−kanamycin medium also containing 100 μM kanamycin. After growing at 37 °C to an OD_{595} of 0.6, the cultures were induced by adding 1 mM IPTG and grown for an additional 2 h. Cells were harvested by centrifuging at 6000*g* for 20 min, washed with a saline solution (170 mM NaCl), and stored at −80 °C overnight. The cell pellet (5.1 g) was resuspended in 50 mL of lysis buffer (20 mM Tris-HCl, pH 7.0, 5 mM EDTA, 100 mM NaCl, 0.1 mM DTT), and lysozyme was added to a final concentration of 0.2 mg/mL. To inhibit protease activity, phenylmethylsulfonyl fluoride was also added to a final concentration of 0.4 mM. The NaCl concentration was raised to 500 mM, and the mixture was frozen and thawed twice using a dry ice−ethanol bath. The resulting extract was homogenized on ice and centrifuged at 14000*g* for 30 min. The resulting supernatant was Fraction I (56 mL).

A solution of 30% w/v poly(ethylene glycol) 4000 was added dropwise to Fraction I to a final concentration of 8.6% w/v. The precipitate was collected by centrifuging at 13000*g* for 15 min and dissolved in 25 mL of buffer A (50 mM HEPES, pH 7.5, 1 mM EDTA, 0.5 mM DTT) (Fraction II).

Fraction II (25 mL) was loaded onto a 12 cm^2 × 8 cm DEAE-Sepharose Fast Flow column and washed with 100 mL of buffer A containing 50 mM NaCl. The column was eluted stepwise with buffer A containing NaCl at concentrations of 100 mM (220 mL), 200 mM (340 mL), 300 mM (340 mL), and 500 mM (425 mL). The fractions containing SP6 gp74 were combined, and solid ammonium sulfate was added to 80% saturation (51.6 g/mL). The precipitate was collected by centrifuging at 10000*g* for 25 min, dissolved in 10 mL of buffer B (20 mM KPO_4, pH 6.8, 1 mM EDTA, 1 mM DTT), and dialyzed against buffer B (Fraction III).

Fraction III (14 mL) was loaded onto a 0.2 cm^2 × 10 cm phosphocellulose column and washed with 5 mL of buffer B containing 20 mM KCl. The column was then eluted with a 30 mL gradient of buffer B from 20 mM KCl to 600 mM KCl. The fractions containing the SP6 gp74 protein were combined (19.5 mL), and ammonium sulfate was added with stirring to 80% saturation. The precipitate was collected by centrifuging, dissolved in 2 mL of buffer A containing 50 mM NaCl, and dialyzed against the same buffer (Fraction IV).

A small amount of Fraction IV (0.2 mL) was loaded onto a Mono Q HR 5/5 FPLC column, and eluted with a gradient from 100 mM NaCl to 500 mM NaCl. SP6 gp74 eluted at approximately 300 mM NaCl (6.1 mL). The fractions containing gp74 were combined and dialyzed against buffer A containing 50% (v/v) glycerol for storage (Fraction V).

Oligoribonucleotide Synthesis Assay. Oligoribonucleotide synthesis assays were performed essentially as described (*32*). Reactions (10 μL) contained 40 mM Tris-HCl, pH 7.5, 50 mM potassium glutamate, 10 mM $MgCl_2$, 10 mM DTT, 50 μg/mL bovine serum albumin, 0.3 mM each of GTP, ATP, UTP, and CTP, 25 μg/mL M13 or 100 μg/mL ΦX174 ssDNA, and varying concentrations of SP6 gp74. [α-^{32}P]-NTPs or [γ-^{32}P]GTP were added as indicated. After incubation at 37 °C for 60 min, the reactions were stopped by the addition of 5 μL of stop solution (98% formamide, 10 mM EDTA, pH 8.0, 0.1% xylene cyanol FF, and 0.1% bromophenol blue). The products of the reactions were separated by electrophoresis through a 25% polyacrylamide gel containing 3 M urea, and visualized by autoradiography. To identify the first and second nucleotides of the reaction products, standard oligoribonucleotide synthesis reactions were heated in boiling water for 5 min prior to the addition of stop solution and subjected to dephosphorylation by bacterial alkaline phosphatase (Amersham). The dephosphorylation reaction contained 50 mM Tris-HCl, pH 9.0, and 1 mM $MgCl_2$, and was initiated by the addition of 0.01 unit/μL bacterial alkaline phosphatase. The dephosphorylated oligoribonucleotides were separated on an 8% polyacrylamide gel containing 6 M urea. Products were visualized by autoradiography and their amounts measured with a Fuji BAS-1000 phosphorimage analyzer. Assays on synthetic templates used oligonucleotides of the sequence 5′-$(T)_{15}$-GCX-$(T)_{15}$-3′ where X was one of the four deoxyribonucleotides.

RNA-Primed DNA Synthesis Assay. The ability of SP6 gp74 to prime DNA synthesis by the T7 DNA polymerase was measured using M13 ssDNA templates as described (*33*). Reactions (20 μL) contained 40 mM Tris-HCl, pH 7.5, 50 mM potassium glutamate, 10 mM $MgCl_2$, 10 mM DTT, 50 μg/mL bovine serum albumin, 0.3 mM each of GTP, ATP, UTP, CTP, dGTP, dATP, dTTP, and [α-^{32}P]dCTP (70 Ci/mol), 12.5 μg/mL M13 ssDNA, 100 nM T7 DNA polymerase, and varying concentrations of SP6 gp74. The reactions were incubated at 37 °C for 20 min, stopped by the addition of 200 mM EDTA, pH 8.0, to a final concentration of 25 mM, and spotted onto Whatman DE81 filters. The filters were then washed 3 times with 0.3 M ammonium formate and once with 95% ethanol to remove unpolymerized nucleotides. The amount of [^{32}P]DNA retained by the filters was measured with a Beckman LS 6500 scintillation counter.

NTP Hydrolysis Assays. NTP hydrolysis by SP6 gp74 was measured using both radioactive (*34*) and colorimetric assays (*35*). Reactions (50 μL) contained 50 mM Tris-HCl, pH 7.5, 40 mM potassium glutamate, 10 mM $MgCl_2$, 10 mM DTT, equal concentrations of selected ribo- or deoxyribo-NTPs totaling 2.0 mM for all NTPs in solution, 135 nM SP6 gp74, and one of the following types of DNA: M13 ssDNA (17 μg/mL), ΦX174 ssDNA (67 μg/mL), a 30-base single-stranded synthetic oligonucleotide (5′-GCGCG CATAT GCGTG ACCCT AAAGT TATCC-3′, 8.0 μM), salmon sperm DNA (40 μg/mL), or λ phage DNA (2 μg/mL) digested with the *Hin*dIII restriction enzyme. In the radioactive assay, [α-^{32}P]NTPs (5000 Ci/mmol) were incubated at 30 °C for 5 min and stopped by the addition of 150 mM EDTA to a final concentration of 15 mM, and 1 μL of each reaction was spotted on a Baker-flex (Cellulose PEI-F) thin-layer chromatography sheet. After drying, the sheet was prerun in water, developed in a solution containing 1 M formic acid and 0.8 M lithium chloride, dried, and subjected to autoradiography. For the colorimetric assay, free orthophos-

phate (P_i) released from NTP was determined using the method of Ames and Dubin (*35*). Reactions were incubated at 37 °C for 15 min, stopped with 50 μL of a mixture of Norit (16%) and perchloric acid (1.4%), and centrifuged. Water (250 μL) and 700 μL of a mixture containing ascorbic acid (1.4%) and ammonium molybdate (0.36%) in 1 N H_2SO_4 were added to 50 μL of each reaction. The resulting solutions were incubated at 42 °C for 20 min to develop a blue color. In this assay, 50 nmol of orthophosphate (P_i) yielded an A_{780} of 1.0.

Helicase Assay. The ability of SP6 gp74 to unwind DNA was measured using a substrate consisting of circular M13 ssDNA and an annealed oligonucleotide possessing an 11-nucleotide 3′-noncomplementary tail (*28, 36*). To prepare this substrate, a 37-nucleotide oligomer (5′-TCACG ACGTT GTAAA ACGAC GGCCA GTTTT TTTTT TT-3′) was radiolabeled at its 3′-end using T4 polynucleotide kinase. The kinase reaction (14 μL) contained 4 pmol of the oligonucleotide in 41 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 143 mM NaCl, 0.36 μg/μL bovine serum albumin, 0.5 μM [γ-^{32}P]ATP (6000 Ci/mmol), and 1 unit/μL T4 polynucleotide kinase and was performed for 15 min at 37 °C. The reaction mixture was heated to 70 °C for 15 min to inactivate the kinase.

M13 ssDNA (5 μg) was added to the [5′-^{32}P]oligonucleotide in 94 mM NaCl and 8 mM $MgCl_2$. The annealing reaction was heated to 65 °C and cooled to room temperature over the course of 60 min. The annealed substrate was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (24:24:1) and purified from [γ-^{32}P]ATP using a 1 mL Sepharose CL-6B column with an eluant of 20 mM Tris-HCl, pH 7.5, 2 mM EDTA, and 100 mM NaCl. The annealed substrate eluted in the void volume.

The helicase reactions (10 μL) contained 40 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 50 μg/mL bovine serum albumin, equal concentrations of selected NTPs (GTP, ATP, UTP, CTP) or dNTPs (dGTP, dATP, dTTP, or dCTP) totaling 2 mM, 1 μL of the helicase substrate, and varying concentrations of the SP6 gp74. The reactions were incubated at 30 °C and stopped by the addition of 0.5 M EDTA to a final concentration of 45 mM. The products were separated on a 10% polyacrylamide Tris-HCl Ready Gel (Bio-Rad) with 0.5× TBE (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA) and visualized by autoradiography or phosphorimage analysis.

RESULTS

Cloning and Sequencing of the SP6 Primase Gene. Preliminary analysis of the SP6 genome[3] revealed the presence of an open reading frame similar to the T7 gene 4 helicase/primase protein. This putative helicase/primase gene encodes a 661 amino acid protein with a mass of 74 kDa (gp74). The amino acid sequence of SP6 gp74 is 22.4% identical to the T7 gene 4 primase/helicase protein (GeneStream align at http://vega.igh.cnrs.fr/bin/align-guess.cgi). The open reading frame of interest was amplified in a polymerase chain reaction using oligonucleotide primers containing restriction sites for the enzymes *Nco*I and *Bam*HI. The PCR product was purified, digested with the *Nco*I and *Bam*HI restriction enzymes, and ligated into similarly treated pET24d vector to produce a plasmid (pETgp74) that expressed a 74-kDa protein. To check for mutations in the expressed protein, plasmid pETgp74 was isolated from HMS174(DE3) cells and sequenced. The sequence of the recombinant putative SP6 helicase/primase gene was identical to the genomic sequence (GenBank accession no. AF159357).

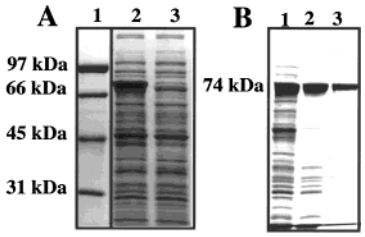

FIGURE 1: Expression and purification of SP6 gp74. Panels A and B show 10% SDS−polyacrylamide Tris-HCl Ready Gels (BioRad) stained with Coomassie Brilliant Blue. (A) Expression of SP6 gp74 under the control of a T7 promoter. Lane 1 contains molecular weight markers. Lane 2 shows a soluble lysate (20 μg) of induced cells containing the pETgp74 plasmid. Lane 3 shows a soluble lysate (20 μg) from induced cells harboring the pETgp74$_{frameshift}$ plasmid. (B) Purification of SP6 gp74. Lane 1, 20 μg of crude lysate (Fraction I). Lane 2, 10 μg of Fraction II. Lane 3, 5 μg of DEAE Sepharose Fast Flow column pool (Fraction III).

Expression and Purification of SP6 gp74. When induced with IPTG, *E. coli* HMS174(DE3) cells containing the plasmid pETgp74 expressed high levels of a protein with an apparent molecular weight of approximately 74 000 (Figure 1A, lane 2). To examine if the expressed protein originated from the first ATG codon in the open reading frame or a downstream start codon, a second plasmid was constructed. The second plasmid, pETgp74$_{frameshift}$, was identical to the pETgp74 plasmid except for a 4-nt (ATGG) insert before the start site of the SP6 gene that placed the SP6 gene out of frame with the T7 promoter. Indeed, cells containing pETgp74$_{frameshift}$ did not express the 74-kDa protein (Figure 1A, lane 3), suggesting that in cells containing the pETgp74 plasmid, gp74 is synthesized from the translational start site identified by preliminary sequencing. In contrast to T7 gene 4, which encodes two proteins from separate in-frame translational start sites, the SP6 gene apparently encodes a single polypeptide. The amount of SP6 gp74 expressed by this system was 10-fold greater than that of similar pET24 constructs expressing the 63-kDa gene 4 primase.[4] Most of the gp74 protein was soluble after lysis of the cells. Such a high level of expression relative to the T7 gene 4 protein suggests that the SP6 protein is either less toxic or more stable when expressed in *E. coli* than is the T7 primase/helicase. The purification of SP6 gp74 was followed by SDS−polyacrylamide gel electrophoresis prior to the discovery of its activity. SP6 gp74 was the predominant protein in extracts of induced cultures of HMS174(DE3) containing plasmid pETgp74 (Figure 1B, lane 1). SP6 gp74 precipitated with nucleic acids in the presence of poly(ethylene glycol) (Figure 1B, lane 2), suggesting that the protein binds DNA or RNA. The protein is more than 90% pure after separation from nucleic acids using a DEAE-Sepharose Fast Flow column (Figure 1B, lane 3). Phosphocellulose and Mono Q columns were used to remove remaining contaminants. Approximately 20% of the amount of gp74 present in the crude extract was recovered after chromatography on phosphocellulose. The described puri-

[4] D. N. Frick, S. Tabor, and C. C. Richardson, unpublished data.

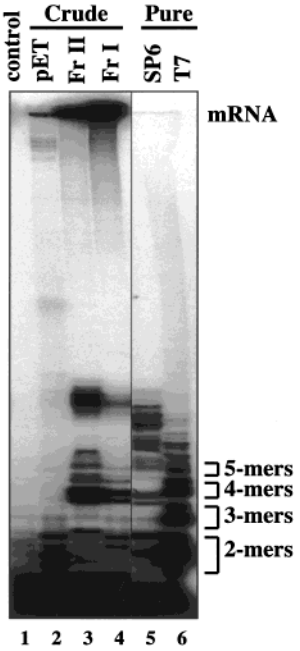

FIGURE 2: Oligoribonucleotide synthesis by SP6 gp74. Oligoribonucleotide synthesis reactions containing GTP, ATP, UTP, [α-^{32}P]-CTP, and M13 ssDNA were performed with no protein (lane 1), 16 μg/mL crude extract from HMS174(DE3) cells containing pET24d (lane 2), 16 μg/mL SP6 gp74 Fraction II (lane 3), 354 μg/mL Fraction I (lane 4), 215 nM SP6 gp74 Fraction V (lane 5), and 500 nM 63-kDa T7 primase/helicase (lane 6). The products of the reactions were separated on a 25% polyacrylamide gel in 3 M urea and detected by autoradiography. The bands at the top of the gel are likely mRNA. Primase-synthesized oligoribonucleotides migrate near the bottom of the gel. The distance from the wells at the top of the gel to the [α-^{32}P]CTP at the bottom of the gel is 27 cm.

fication yields 1.2 mg of pure gp74 protein per liter of cell culture.

SP6 gp74 Catalyzes DNA-Dependent Oligoribonucleotide Synthesis. Primases provide primers to DNA polymerases to initiate DNA synthesis. Generally these primers are short oligoribonucleotides (*1*). In the presence of NTPs and a ssDNA template, a DNA primase synthesizes these oligoribonucleotides. Whereas other polymerases will use any sequence as a template, bacterial and phage DNA primases ordinarily synthesize oligoribonucleotides only at certain DNA sequences. For example, with ΦX174 (*3*) or M13 (*37*) ssDNA as a template, the T7 primase synthesizes primers at primase recognition sites containing the sequences 5′-(G/T)GGTC-3′ or 5′-GTGTC-3′. All of these primase recognition sites begin with the trinucleotide 5′-GTC-3′. This trinucleotide is necessary and sufficient to support dinucleotide synthesis by the T7 primase (*4*). Oligoribonucleotide synthesis begins opposite the dT in the recognition site, and all primers retain the triphosphate moiety of ATP at the 5′-terminus (*5, 6*). The essential 3′-dC in the recognition site is not copied into the primer and is designated as "cryptic" (*3, 7*). Because of the potential for the SP6 primase to exhibit similar template requirements, oligoribonucleotide synthesis by gp74 was first examined using long heterogeneous ssDNA templates.

Oligoribonucleotide synthesis activity was observed in crude extract of cells expressing SP6 gp74, and this activity remained throughout the purification. Figure 2 shows oligoribonucleotide synthesis reactions performed with no protein (lane 1), a crude extract from HMS174(DE3) cells not expressing gp74 (lane 2), SP6 gp74 Fraction II (lane 3), SP6 gp74 Fraction I (lane 4), SP6 gp74 Fraction V (lane 5), and T7 primase (lane 6). All reactions contained M13 ssDNA, GTP, ATP, UTP, [α-^{32}P]CTP, and magnesium. In the presence of M13 ssDNA, purified gp74 (lane 5), as well as the crude extract of cells expressing SP6 gp74 (lane 4), catalyzed the synthesis of short oligoribonucleotides similar in size to those produced by the purified T7 gene 4 primase (lane 6). Extracts of cells containing the pET24d vector alone did not (lane 2). The amount of oligoribonucleotide synthesis catalyzed by purified SP6 gp74 (Fraction V) was proportional to both reaction time and enzyme concentration and was dependent upon the addition of ssDNA. The synthesis of a variety of oligoribonucleotides was detected at 20 min, with the amount of oligoribonucleotide synthesis increasing up to 120 min, indicating that gp74 acts catalytically. Other ssDNA templates also support oligoribonucleotide synthesis. With ΦX174 ssDNA, oligoribonucleotide synthesis could be observed with as little as 3 nM SP6 gp74, and synthesis was proportional to enzyme concentration.

As with other known RNA polymerases, the oligoribonucleotide synthesis catalyzed by SP6 gp74 requires magnesium, and the optimal pH is approximately 7.5. The addition of 50−100 mM potassium glutamate, which enhances the primase activity of the T7 gene 4 protein (*38*), increased the rate of oligoribonucleotide synthesis by SP6 gp74 approximately 10%. The presence or absence of the reducing agent DTT did not affect the activity of the SP6 enzyme.

Because gp74 was purified from *E. coli* bacteria using a T7 expression system, there was the potential for contamination with both the *E. coli* and T7 RNA polymerases. T7 RNA polymerase was likely separated from gp74 during purification. In contrast to gp74, which elutes from DEAE-Sepharose at a relatively high salt concentration of 300 mM NaCl, T7 RNA polymerase does not bind to DEAE resins at this ionic strength (*39*). None of the subunits of the *E. coli* RNA polymerase holoenzyme (36.5 kDa, 151 kDa, and 156 kDa) were observed in the final fractions of the purification. To ensure that a small amount of *E. coli* RNA polymerase was not responsible for the oligoribonucleotide synthesis observed in the SP6 gp74 preparation, RNA-primed DNA synthesis assays (see below) were performed with rifampicin, an antibiotic known to selectively inhibit *E. coli* RNA polymerase, with 50% inhibition at a concentration of 0.01 μg/mL rifampicin (*40*). RNA-primed DNA synthesis reactions with SP6 gp74 or the 63-kDa T7 gene 4 protein were performed in the presence of 10 μg/mL and 100 μg/mL rifampicin. Neither gp74 nor T7 gene 4 protein was inhibited. The analysis of the reaction products (Figure 2) also provides strong support that the purified gp74 was not contaminated with other DNA-dependent RNA polymerases. Crude extracts of cells containing either pETgp74 or pET24d catalyzed the synthesis of long RNA products that did not enter the gel matrix, while purified gp74 did not. Such products are likely RNAs synthesized by the T7 or *E. coli* RNA polymerases. The synthesis of such long RNA products is not catalyzed by purified gp74, indicating that gp74 was separated from contaminating RNA polymerases.

SP6 gp74 Primes DNA Synthesis. In vivo, primases catalyze the synthesis of the oligoribonucleotides used by

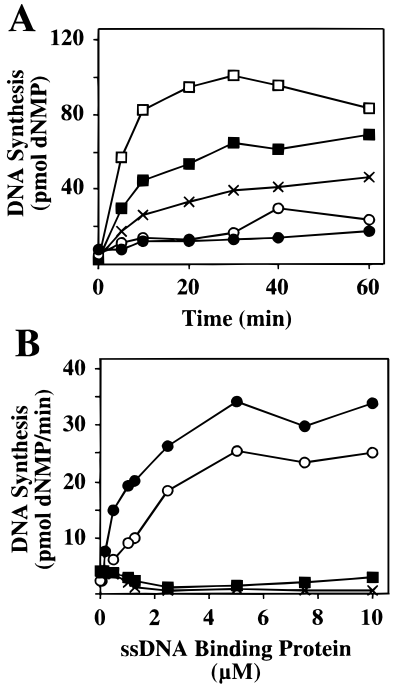

FIGURE 3: RNA-primed DNA synthesis supported by SP6 gp74. RNA-primed DNA synthesis was measured in the presence of 12.5 μg/mL M13 ssDNA template, GTP, ATP, UTP, CTP, dGTP, dATP, dTTP, and [α-^{32}P]dCTP. (A) Reactions were performed with 100 nM T7 DNA polymerase and 34 nM (●), 68 nM (○), 135 nM (×), and 270 nM (■) SP6 gp74. As a positive control, a reaction was performed with 13 nM 63-kDa T7 gene 4 primase/helicase (□). Reactions were incubated at 37 °C, and samples were withdrawn at 0, 5, 10, 20, 30, 40, and 60 min and spotted on DE81 filters. [^{32}P]DNA was measured by counting the radioactivity remaining on the filters after washing. (B) RNA-primed DNA synthesis reactions containing 100 nM T7 DNA polymerase, 12.5 μg/mL M13 ssDNA template, GTP, ATP, UTP, CTP, dGTP, dATP, dTTP, [α-^{32}P]dCTP, and either 135 nM SP6 gp74 or 13 nM 63-kDa T7 gene 4 protein were titrated with *E. coli* and T7 phage ssDNA binding protein as follows: T7 gene 4 with *E. coli* ssDNA binding protein (●), T7 gene 4 with T7 gp2.5 (○), SP6 gp74 with *E. coli* ssDNA binding protein (×), and SP6 gp74 with T7 gp2.5 (■).

DNA polymerases to prime DNA replication. To test for such an activity with purified SP6 gp74, RNA-primed DNA synthesis reactions were performed in which the amount of radiolabeled nucleotides incorporated into DNA by a DNA polymerase was measured in the presence and absence of a DNA primase. Because SP6 DNA polymerase was unavailable for this study, T7 DNA polymerase was used instead. The experiment in Figure 3A shows the amount of DNA synthesis in reactions containing 34−270 nM SP6 gp74. T7 DNA polymerase was present in all reactions at a concentration of 100 nM. The initial rate of DNA synthesis increased with increasing concentrations of the SP6 protein. Despite the use of 20-fold higher protein concentrations, the SP6 primase stimulated DNA synthesis by the T7 polymerase less efficiently than the T7 primase. Perhaps this is attributable to a protein−protein interaction between the T7 primase and T7 polymerase that is not satisfied by the SP6 primase. Nonetheless, the oligoribonucleotides synthesized by SP6 primase were fully functional as primers for DNA polymerase.

DNA polymerase is likely but one of several proteins at the replication fork with which the primase must interact during DNA replication. For example, the *E. coli* DnaG primase switches (*41*) between essential contacts with the polymerase and the *E. coli* ssDNA binding protein to efficiently prime DNA synthesis (*42*). Likewise in the T7 replisome, both *E. coli* and T7 gene 2.5 ssDNA binding proteins increase the rate of initiation of lagging strand DNA synthesis by the T7 primase up to 100-fold (*43*) and stimulate primer synthesis by at least 5-fold (*4*). To determine if ssDNA binding proteins enhance the activity of the SP6 primase in a similar fashion, RNA-primed DNA synthesis assays were carried out with varying concentrations of *E. coli* and T7 gp2.5 ssDNA binding proteins (Figure 3B). Both ssDNA binding proteins stimulated DNA synthesis with the T7 primase severalfold. The opposite effect was observed with the SP6 primase. Less than 25% of the original activity with SP6 gp74 remained after the addition of 2.5 μM of either *E. coli* or T7 gp2.5 ssDNA binding proteins, indicating a possible inhibitory effect of these two proteins. These data support the notion that the T7 primase and ssDNA binding proteins work together (*43*) at the replication fork and highlight another important difference between the T7 and SP6 primases. Perhaps the SP6 primase requires the presence of *S. typhimurium* ssDNA binding protein or other SP6 replication proteins for optimal primer synthesis.

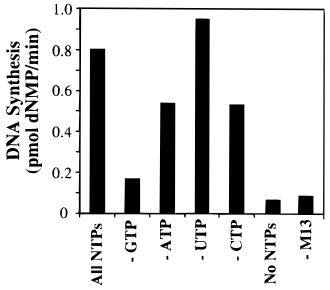

FIGURE 4: Ribonucleoside triphosphate requirements for priming of DNA synthesis by the SP6 primase. RNA-primed DNA synthesis was measured using 135 nM SP6 gp74, 100 nM T7 DNA polymerase, and 12.5 μg/mL M13 ssDNA template. Standard reactions contained 0.3 mM each of GTP, ATP, UTP, CTP, dGTP, dATP, dTTP, and [α-^{32}P]dCTP. Individual NTPs were omitted from the reaction as indicated. Also shown is a reaction with all NTPs but lacking M13 ssDNA template.

The SP6 Primase Initiates Oligoribonucleotides with pppGpC. RNA-primed DNA synthesis reactions were performed with various combinations of NTPs in order to determine the nucleotide composition of the primers (Figure 4). When GTP was omitted, RNA-primed DNA synthesis was reduced to background levels. GTP is thus essential for the priming of DNA synthesis by SP6 primase. The SP6 primase shows a less stringent requirement for ATP and CTP, with DNA synthesis reduced by 40% in the absence of either NTP. SP6 primase does not require UTP. When no ribonucleoside triphosphates are included in the reaction, no DNA synthesis is observed above background levels, suggesting that dNTPs cannot be efficiently used for primer synthesis. These data suggest that gp74 synthesizes oligoribonucleotides containing GMP, AMP, or CMP.

The composition of the oligoribonucleotide products of the SP6 primase was directly determined by performing four sets of oligoribonucleotide synthesis assays using M13 ssDNA and different combinations of the four NTPs (Figure 5, A−D). With each set of reactions, various combinations of unlabeled NTPs were included with a single [α-^{32}P]NTP to generate unique sets of radioactive oligoribonucleotide

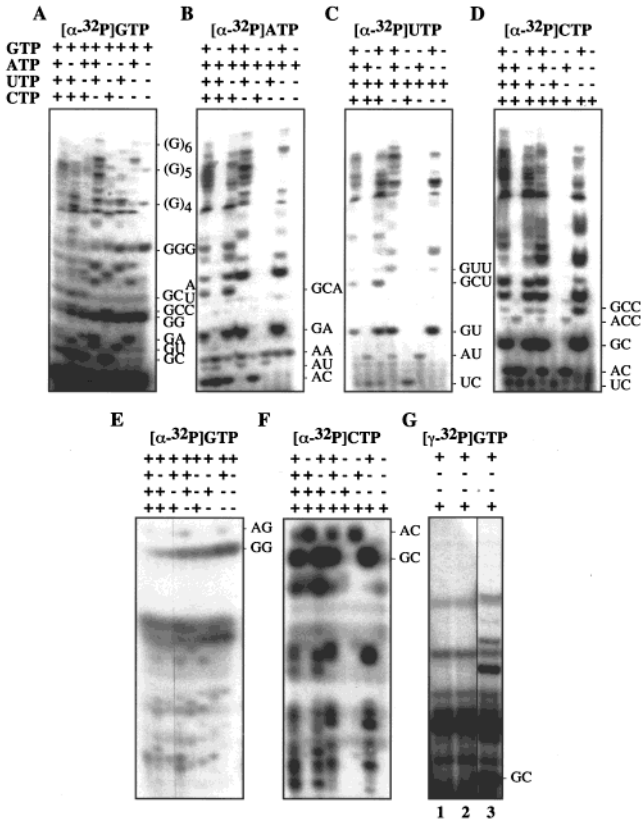

FIGURE 5: Analysis of oligoribonucleotide products of the SP6 primase. Oligoribonucleotide synthesis reactions in panels A−F contained 135 nM SP6 gp74 and 25 μg/mL M13 ssDNA template. (A−D) Oligoribonucleotide synthesis reactions were performed in the presence of the NTPs and [α-^{32}P]NTPs indicated above each panel. Products were separated on a 25% polyacrylamide gel in 3 M urea. An autoradiogram of the gel is shown. (E and F) The reactions in panels A and D were dephosphorylated with bacterial alkaline phosphatase and separated using an 8% polyacrylamide gel in 6 M urea. An autoradiogram of the gel is shown. Note that shorter (<6 nt) dephosphorylated oligoribonucleotides migrate more slowly than longer products. (G) Oligoribonucleotide synthesis reactions using 68 nM SP6 gp74 were performed in the presence of [γ-^{32}P]GTP and CTP. Lane 1 shows the products of a reaction containing no DNA template. The reaction in lane 2 contained no enzyme. In lane 3, a 2.0 μM sample of the *Bam*HI-2 oligonucleotide (see Experimental Procedures for sequence) was used as a DNA template. The reaction products were analyzed as in panels A−D. [γ-^{32}P]GTP exhibits a high radioactive background present in all three lanes.

products. No oligoribonucleotide synthesis was detected in control reactions lacking SP6 primase or DNA template, or if [α-^{32}P]NTP's were substituted with either [α-^{32}P]dGTP, [α-^{32}P]dATP, [α-^{32}P]dTTP, or [α-^{32}P]dCTP (data not shown). Using M13 ssDNA as a template, the SP6 primase generates a myriad of oligoribonucleotides. The rates of migration of these products compare closely to the well-characterized 2−5-nucleotide-long oligoribonucleotides synthesized by the T7 primase (Figure 2, lanes 1 and 2). When GTP was omitted from any reaction in Figure 5, the synthesis of longer oligoribonucleotides (>3 nt) was eliminated, confirming that the SP6 primase has a stringent requirement for GTP. Reactions containing only [α-^{32}P]GTP (Figure 5A) produced a ladder of products that correspond to guanine homopolymers ranging in length from two to six nucleotides. By including ATP, UTP, or CTP separately, all four dinucleotides containing guanine can be identified. The relative rate of migration of these four dinucleotides is as follows: GC > GU > GA > GG. By further examining the oligoribonucleotide species generated by the inclusion of various NTP combinations, the compositions of the other dinucleotides and several trinucleotides may be deduced (Figure 5, panels A−D). The dinucleotide composed of guanosine and cytidine was the major product observed with [α-^{32}P]GTP (Figure 5A) when all four NTPs or GTP and CTP were present in the reactions. Likewise, in reactions containing [α-^{32}P]CTP, the major radioactive product was also GC (Figure 5D). In all reactions containing both GTP and CTP, the primary radioactive dinucleotide product was also GC. In reactions containing [α-^{32}P]ATP (Figure 5B), the major radioactive dinucleotides produced in the presence of all four NTPs were of the compositions AC and GA. The GC dinucleotides would not be visible in reactions using [α-^{32}P]ATP. However, in reactions containing [α-^{32}P]CTP or [α-^{32}P]GTP, more GC is synthesized than AC or GA, respectively. In reactions containing [α-^{32}P]GTP, it is clear that more of the dinucleotide GC is synthesized than GA (Figure 5A). Likewise, in reactions containing [α-^{32}P]CTP, more of the dinucleotide GC is synthesized than AC (Figure 5D). Reactions with [α-^{32}P]UTP (Figure 5C) produced fewer radioactive products in lower quantities, confirming that UMP is incorporated infrequently into oligoribonucleotide products. After adjusting for the specific activity of each [α-^{32}P]NTP, we conclude that the major dinucleotide product synthesized by the SP6 primase is either 5′-GC-3′ or 5′-CG-3′. However, the experiments shown in Figure 5, panels A−D, could not distinguish between these two sequences.

To determine whether the nucleotide at the 5′-end of the oligoribonucleotide products is CTP or GTP, the oligonucleotides shown in Figure 5, panels A and D, were dephosphorylated with bacterial alkaline phosphatase and reanalyzed (Figure 5, panels E and F). Bacterial alkaline phosphatase hydrolyzes terminal phosphates while leaving internal phosphodiester bonds intact. After removal of the 5′-phosphates, the relative mobility of short oligoribonucleotides is altered (*44*). After dephosphorylation, the dinucleotides migrate at the slowest rate, appearing near the top of the gel. In reactions containing [α-^{32}P]CTP, the dinucleotide products remain labeled after dephosphorylation with bacterial alkaline phosphatase, indicating that the labeled phosphate of CTP must be an internal phosphate group. Therefore, the CTP is not incorporated at the 5′-end of the oligoribonucleotides and must be incorporated as the second nucleotide. Dephosphorylated products of [α-^{32}P]GTP reactions show that guanosine is incorporated at the 5′-end of the dinucleotide product, as the GC product disappears with phosphatase treatment. Furthermore, the products of oligoribonucleotide synthesis reactions performed with [γ-^{32}P]GTP (in the absence of bacterial alkaline phosphatase) include a pppGpC product (Figure 5, panel G), confirming the retention of the 5′-triphosphate moiety of GTP at the 5′-end of the gp74-synthesized oligoribonucleotide product. We conclude that the sequence of the 5′-end of the majority of the oligoribonucleotides synthesized by the SP6 primase is pppGpC.

The SP6 Primase Recognition Site Contains a Cryptic dA at the 3′-End. The DNA primases from *E. coli* and phages T7 and T4 each begin template-directed oligoribonucleotide synthesis at unique trinucleotide recognition sites (*8, 9, 15, 45, 46*). For example, the T7 primase initiates oligoribo-

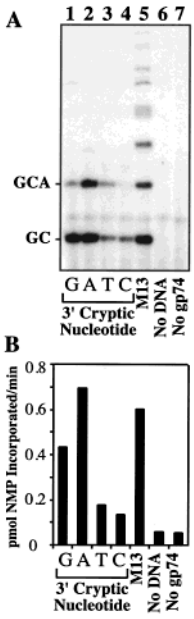

FIGURE 6: SP6 primase catalyzes oligoribonucleotide synthesis on short synthetic DNA templates. Oligoribonucleotide synthesis reactions using 135 nM SP6 gp74 were performed with 2 μM synthetic oligonucleotide DNA templates of the sequence 5′-$(T)_{15}$-GCX-$(T)_{15}$-3′ where the 3′ cryptic nucleotide, X, was one of the four deoxyribonucleotides. The reactions contained 300 μM each of ATP, UTP, GTP, and [α-^{32}P]CTP. Lanes 1−7 in (A) correspond to bars 1−7 in (B). (A) Oligoribonucleotide products were separated on a 25% polyacrylamide gel in 3 M urea and visualized by autoradiography. In lanes 1−4, assays were performed on oligonucleotide templates containing the 3′ cryptic nucleotides G, A, T, and C, respectively. Lane 5 shows the products of oligonucleotide synthesis on an M13 ssDNA template. Lanes 6 and 7 were negative controls in which DNA template and gp74 were omitted, respectively. (B) The amount of each oligoribonucleotide in (A) was measured using phosphorimage analysis and the rate of ribonucleotide incorporation calculated.

nucleotide synthesis at the sequence 5′-GTC-3′ by synthesizing the dinucleotide pppApC (*33*). The 3′-nucleotide is designated as "cryptic" because, while it is required for recognition, its complement is not incorporated into the oligoribonucleotide products. If the SP6 primase begins each primer with the dinucleotide pppGpC, then the complementary DNA template recognized by the protein must include the sequence 5′-GC-3′. If gp74 also requires a cryptic 3′-nucleotide as the aforementioned primases do, then it would synthesize primers at the trinucleotide sequence 5′-GCX-3′ where X is the cryptic nucleotide. To determine if the nucleotide flanking the 3′-end of the DNA sequence encoding the initial pppGpC product influences the rate of primer synthesis by SP6 gp74, oligoribonucleotide synthesis reactions containing GTP and [α-^{32}P]CTP were carried out on defined oligonucleotide templates of the sequence 5′-$(T)_{15}$-GCX-$(T)_{15}$-3′, where only the 3′-nt, X, was varied among the four deoxynucleotides (Figure 6). Oligoribonucleotides were synthesized at the highest rate (0.69 pmol of NMP incorporated/min) by gp74 in the presence of oligonucleotide templates containing the sequence 5′-$(T)_{15}$-GCA-$(T)_{15}$-3′ (Figure 6, lane 2). Oligoribonucleotides were synthesized less rapidly (0.43 pmol/min) in the presence of oligonucleotide templates possessing a guanine at the cryptic position [5′-$(T)_{15}$-GCG-$(T)_{15}$-3′] (lane 1). Templates with dT or dC at the cryptic position supported ribonucleotide incorporation at the slower rates of 0.18 and 0.14 pmol/min, respectively. These data indicate that the preferred cryptic nucleotide is dA and that the strongest SP6 gp74 recognition sequence is 5′-GCA-3′. The requirement for a cryptic dA is less stringent than the requirement of the T7 primase for its recognition sequence (*47*). However, the possibility remains that the SP6 primase requires more than a simple trinucleotide recognition sequence for maximum activity.

The rates of oligoribonucleotide synthesis (Figure 6) also provide direct evidence that gp74 acts catalytically. When the template 5′-$(T)_{15}$-GCA-$(T)_{15}$-3′ was used (Figure 6A, lane 2), 1.35 pmol of enzyme catalyzed the incorporation of 41.6 pmol of ribonucleotides during the 60 min reaction period, indicating an enzyme turnover rate (k_{cat}) of 0.009 s^{-1}. This rate of oligoribonucleotide synthesis is 10-fold faster than that measured for the *E. coli dnaG* primase (0.00089 s^{-1}) on short synthetic DNA templates in the absence of any auxiliary proteins (*46*).

Lack of Apparent Helicase Activity in SP6 gp74. The fact that the amino acid sequence of SP6 gp74 contains the five regions conserved among DnaB-like helicases suggests that it could function—like the T7 63-kDa gene 4 protein—as a helicase as well as a primase. DNA helicases use the energy of nucleotide hydrolysis to unwind duplex DNA molecules and were originally discovered as DNA-dependent ATPases (*1*). Helicase activity may be monitored by measuring DNA-dependent NTP hydrolysis. Assays performed with [γ-^{32}P]-GTP and [γ-^{32}P]ATP using SP6 gp74 and M13 ssDNA did not show evidence of DNA-dependent NTPase activity (data not shown). To test for DNA-dependent NTP hydrolysis with other DNAs and NTPs, a colorimetric assay for orthophosphate release from NTPs was also used (Figure 7A). The low rate of NTP hydrolysis observed in some reactions in Figure 7A was not consistently above background and was not proportional to the amount of SP6 gp74 present in the reactions. DNA unwinding was also measured directly by examining the ability of SP6 gp74 to displace a [5′-^{32}P]-oligonucleotide from circular M13 ssDNA (Figure 7B) in the presence of all eight canonical NTPs. Under these conditions, no significant helicase activity was observed even at a gp74 concentration far exceeding the level of T7 helicase required to unwind all of the DNA.

DISCUSSION

DNA primases play an integral role in the initiation of both leading and lagging strand DNA synthesis. Because of the complexities of the DNA replication systems in eukaryotic organisms, and even single-celled bacteria, much attention has been focused on bacteriophages as models for the understanding of DNA synthesis. Bacteriophage DNA replication complexes offer a number of advantages over the complexes found in other organisms. In the phage T7, for example, the number of proteins involved in DNA replication is dramatically lower than the number of proteins present in the DNA replication complexes of *E. coli* or humans. The only four proteins required for in vitro T7 DNA synthesis are the T7 gene 5 DNA polymerase, its accessory protein *E. coli* thioredoxin, the T7 gene 2.5 ssDNA binding protein, and the T7 gene 4 protein (*18*). The last protein, like proteins from *E. coli* phages T3 and P4, combines primase and helicase activities into a single polypeptide.

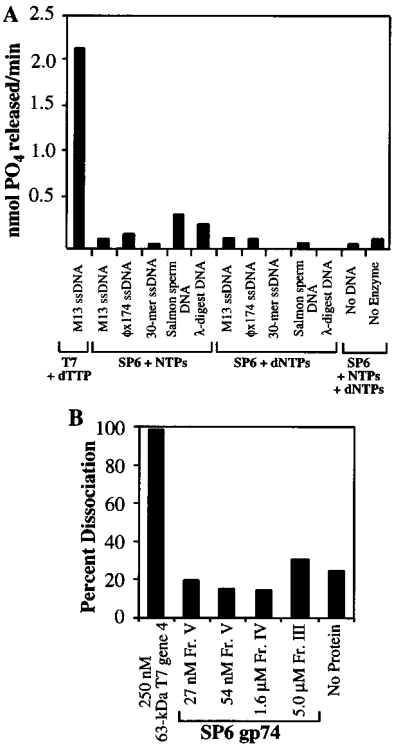

FIGURE 7: SP6 primase lacks detectable NTP hydrolysis and helicase activities. (A) Colorimetric NTP hydrolysis assays were performed using either 135 nM SP6 gp74 or 100 nM 63-kDa T7 gene 4 protein in the presence of the indicated DNA templates. Reactions contained GTP, ATP, CTP, and UTP (NTPs); or dGTP, dCTP, dTTP, and dATP (dNTPs); or all eight NTPs and dNTPs. (B) Helicase assays were performed with the indicated concentrations of 63-kDa T7 gene 4 protein and SP6 gp74 using a substrate consisting of a [5′-^{32}P]oligonucleotide partially annealed to an M13 ssDNA circle. All reactions contained 0.25 mM each of GTP, ATP, UTP, CTP, dGTP, dATP, dTTP, and dCTP. The products of these reactions were separated on a 10% polyacrylamide gel, and the percentage of free oligonucleotide was determined by phosphorimager analysis.

Sequence data indicate that the *S. typhimurium* phage SP6, a distant relative of T7 (*29*−*31*), possesses a gene that encodes a protein homologous to the primase/helicase protein of the T7 and T3 bacteriophages. We have designated the 74-kDa protein product of this gene as gp74 and shown that it is a primase. Like other DNA primases, the SP6 primase exhibits DNA-dependent RNA polymerase activity and synthesizes oligoribonucleotides up to 6 nucleotides in length that can serve as primers for T7 DNA polymerase. The oligoribonucleotides synthesized by SP6 primase are similar in length (2−6 nucleotides) to the products synthesized by the T4 (*15*) and the T7 (*48*) primases. The oligoribonucleotides synthesized by the *E. coli dnaG* primase (*49*) and the eukaryotic primase from calf thymus (*50*) are approximately twice as long as the products of SP6 primase. The SP6 gp74 primase is also similar to other primases in that it requires divalent metal cations and exhibits optimal activity at pH 7.5.

SP6 gp74 is a DNA primase with several novel characteristics. One distinctive property is its substrate specificity. SP6 primase efficiently synthesizes oligonucleotides using only GTP and CTP. The T7 primase, its closest known relative, uses ATP and CTP for primer synthesis (*5*). Our data also suggest that most of the oligoribonucleotides synthesized by the SP6 primase begin with the dinucleotide pppGpC. For the synthesis of this dinucleotide, SP6 gp74 recognizes a trinucleotide template containing a cryptic purine at the 3′-end. The two 5′-nucleotides of this 5′-GCA-3′ recognition sequence serve as the start of the DNA template from which SP6 primase synthesizes oligoribonucleotides. All of the other well-characterized prokaryotic and phage primases recognize different trinucleotide sequences. The *E. coli* DnaG primase recognizes the trinucleotide 5′-CTG-3′ (*8*). Bacteriophage T7 primase synthesizes oligoribonucleotides at the sequence 5′-GTC-3′ (*3, 4*), and the phage T4 gp61 primase recognizes the sequence 5′-G(T/C)T-3′ (*9*). In each case, the trinucleotide sequence contains a cryptic nucleotide at the 3′-end that is not copied into primers (*3, 7*). The 5′-GCA-3′ recognition sequence is another distinguishing characteristic of the SP6 enzyme.

A BLAST search (*51*) using the SP6 primase amino acid sequence reveals that the protein is most similar to the T3 and T7 gene 4 dual-function primase/helicase proteins. Within the *dnaG*-like class of primases, there exists six widely conserved signature sequences, designated by Ilyina et al. as primase boxes 1 through 6 (*26*). Within the *dnaB*-like class of helicases, there are five widely conserved sequences designated as helicase boxes 1, 1a, and 2 through 4. In SP6 gp74 and the T7 gene 4 protein, the conserved primase boxes occur entirely within the N-terminal halves of the proteins while the helicase boxes exist within the C-terminal regions. Hence, the N-terminal portion of such proteins is called the primase domain, and the C-terminal portion is called the helicase domain. Figure 8 shows an alignment of the amino acid sequences of the SP6 and T7 primases that was done using the program Clustal W (*52*). Boxes indicate the widely conserved primase and helicase motifs, and a sequence of basic amino acids shared among RNA polymerases (*53*). The recently described TOPRIN domain conserved among primases, topoisomerases, OLD family nucleases, and RecR proteins (*54*) spans primase boxes 4 through 6. On Figure 8, identical and similar amino acids are marked. Asterisks (*) denote residues that are invariant in SP6 gp74, the 63-kDa T7 gene 4 protein, and other bacterial and bacteriophage primases and helicases. Residues that are invariant among DnaG-like primases and DnaB-like helicases, but are not conserved in the SP6 primase, are indicated by crosses (×). Double asterisks mark the cysteine and histidine residues believed to coordinate the zinc ion of the zinc finger motif in the primase domain. The identical (red) and similar (blue) residues in this alignment point out amino acids that may be essential for primase-catalyzed oligoribonucleotide synthesis.

Most of the invariant residues of the primase and helicase boxes are present in the SP6 protein. Notable exceptions within the primase domain are glycine-85 of box 2, aspartate-110 of box 3, and glutamate-157 of box 4, which we assume are not necessary for oligoribonucleotide synthesis. The functions of these primase boxes are not fully understood, but it has been speculated that boxes 4 and 5 of the primase domain may be involved in the binding of template DNA (*26*). In the T7 gene 4 protein, primase box 1 has been shown to form a zinc binding motif thought to interact with primase recognition sites in template DNA (*24*). We propose a similar function for primase box 1 in the SP6 primase. The T7 and T3 primases possess a Cys-X_2-Cys-X_n-Cys-X_2-Cys zinc

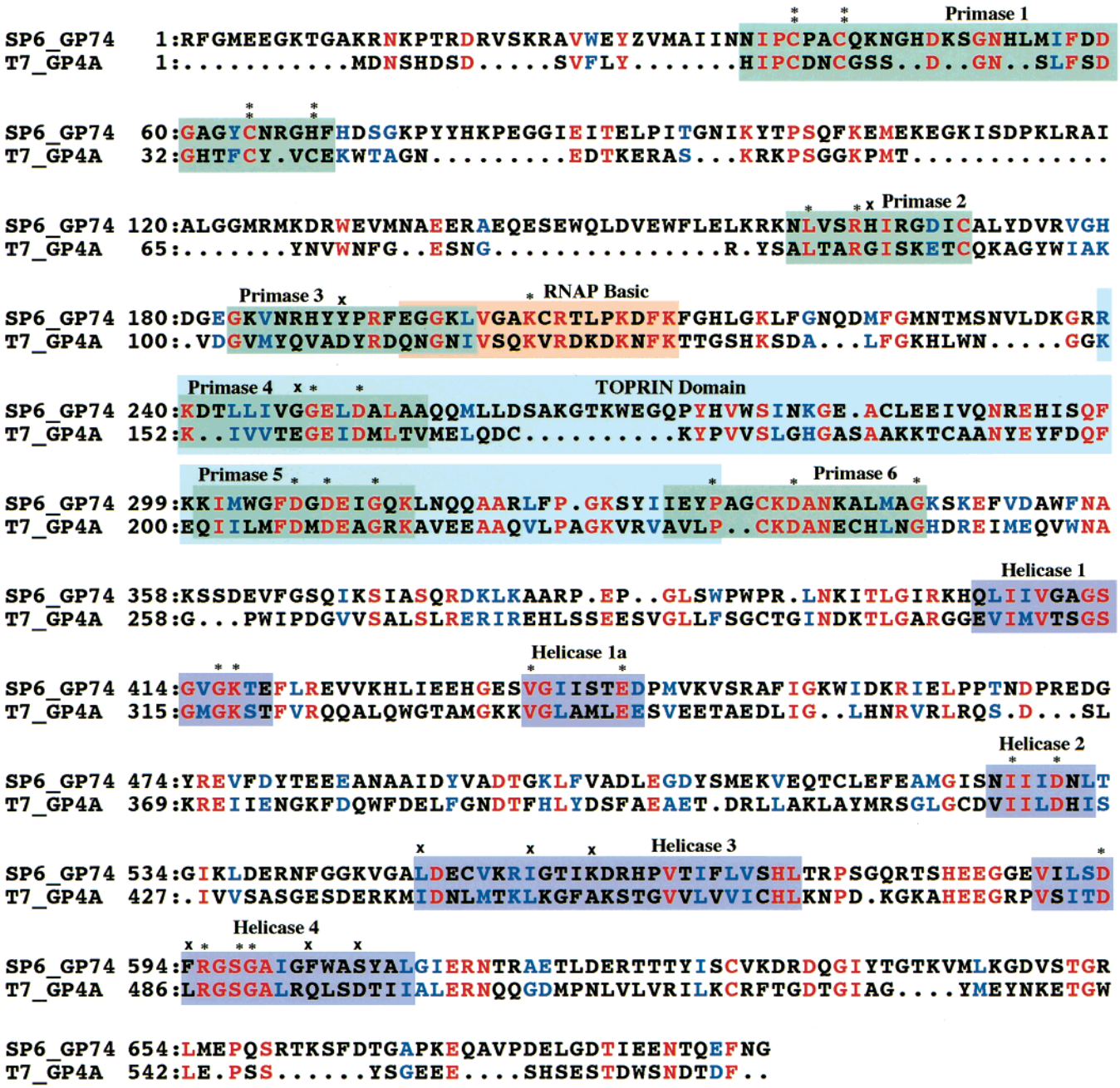

FIGURE 8: Amino acid sequence alignment between SP6 primase and the T7 primase/helicase. The amino acid sequences of the SP6 primase and 63-kDa T7 gene 4 protein were aligned using the program Clustal W (*52*). Boxes indicate the widely conserved primase, helicase (*26*), and RNA polymerase sequences (*53*), and the TOPRIN domain conserved among primases and topoisomerases (*54*). Red type indicates identity whereas blue type indicates similarity. Asterisks (*) denote residues that are invariant among SP6 gp74, T7 gene 4 protein, and other bacterial and bacteriophage primases and helicases. Crosses (×) mark residues that are widely conserved among bacterial and bacteriophage primases and helicases, but are not conserved in SP6 gp74. Double asterisks indicate the cysteine and histidine residues thought to coordinate zinc.

binding motif, and bacterial DnaG-like primases possess a zinc-containing Cys-X_2-His-X_n-Cys-X_2-Cys motif. SP6 primase box 1 contains a different sequence, Cys-X_2-Cys-X_{21}-Cys-X_3-His. The Cys and His residues in this motif could all bind zinc, forming a fingerlike motif. This zinc finger may be involved in the interaction of gp74 with primase recognition sites in DNA, in maintaining the structural integrity of the enzyme, or in interactions with other proteins at the replication fork. Based on the role of this motif in the T7 primase, we prefer the first hypothesis. Amino acid substitutions in box 1 of the T7 primase not only perturb zinc binding of the protein (*55*) but also alter the sequence specificity of the primase (*24*, *56*). When histidine-33 of the 63-kDa T7 gene 4 protein is changed to alanine, the protein no longer synthesizes oligoribonucleotides in the presence of the 3′-CTG-5′ recognition sequence. Instead, the T7 gene 4 H33A protein synthesizes oligoribonucleotides at sequences containing a cryptic purine (i.e., 5′-GTG-3′ or 5′-GTA-3′). Remarkably the proposed recognition site for the SP6 primase also contains a cryptic purine, and the amino acid that aligns with His-33 of T7 gene 4 protein is alanine (Figure 8).

One of the most pertinent issues for future work on SP6 gp74 is the characterization of the protein's potential helicase

activity, or lack thereof. To this end, gp74 should be assayed on more diverse types of DNA and RNA substrates containing a variety of structures such as 5′ overhangs or synthetic oligonucleotide "forks" that mimic aspects of the DNA replication fork. The homology between the helicase boxes of SP6 gp74 and the T7 gene 4 protein is higher than that for the primase boxes (64.9% vs 53.2% similarity), yet the SP6 protein lacks detectable helicase activity. The nucleotide binding site (*57*) appears to be intact (helicase box 1), and the homology for the rest of the helicase boxes is strong. Besides the T3 and T7 primase/helicase proteins, BLAST searches using the gp74 amino acid sequence returned *only* various replicative DNA helicases. Thus, it is somewhat surprising that gp74 exhibits neither NTP hydrolysis nor helicase activity. A possible explanation may lie with the alteration in SP6 gp74 of 6 of the 16 residues that are invariant among helicases similar to the T7 and bacterial DnaB-like helicases (Figure 8). Although gp74 apparently does not hydrolyze NTPs in a DNA-dependent manner nor unwind DNA (Figure 7), some of these helicase motifs could be used by the protein to bind DNA or NTP substrates. Site-directed mutagenesis of the conserved residues in the SP6 primase should help determine the roles of the helicase motifs in primase function.

The inability to detect helicase activity in gp74 may also be an artifact of the purification process or assay conditions. For example, the T7 gene 4 protein functions as a hexamer (*58, 59*). With SP6 gp74, it is possible that such multiple subunit structures were disrupted during purification and, unlike the T7 helicase/primase, do not properly reassemble in vitro. Three preliminary observations regarding the oligomeric state of gp74 support this contention. First, the gp74 protein elutes from gel filtration columns with proteins that range in mass from 50 to 150 kDa. Second, under the conditions that the T7 gene 4 protein forms hexamers during nondenaturing polyacrylamide gel electrophoresis (*59*), the SP6 protein forms mainly lower molecular weight species. Third, no change in the migration rate of gp74 through native gels was detected in the presence or absence of ATP, Mg^{2+}, or DNA.[5] However, these nonequilibrium methods do not rule out an oligomeric structure for the SP6 primase because subtle variations in experimental conditions have been shown to greatly affect the formation of stable oligomers of the T7 gene 4 primase (*58*). More detailed biophysical analyses will be necessary to determine the quaternary structure of the SP6 primase. The lack of potential SP6 gp74 helicase may, alternatively, result from a requirement of other SP6 or *S. typhimurium* DNA replication associated proteins such as DNA polymerase or ssDNA binding protein.

SP6 gp74 is expressed remarkably well in *E. coli*, suggesting that it is much less toxic to the bacterium than its T7 gene 4 counterpart, which cannot be overexpressed in *E. coli* without a mechanism of inhibiting its lethal dTTPase activity (*59*). A high level of SP6 primase expression could facilitate biophysical and structural analyses. The in vivo role of the SP6 primase in phage replication could also be determined through genetic studies. For example, although the presence of the SP6 primase gene does not permit the growth of T7 phage lacking gene 4,[2] it may complement the temperature-sensitive phenotypes of other primase mutants. The construction of SP6 phage lacking a complete SP6 primase gene will be essential to completely understand the biological role of gp74. Furthermore, now that the properties of SP6 primase have been described, the pg74 could be purified from phage-infected *S. typhimurium* cells to identify potential cofactors and replication proteins that may enhance of its activity. Additional comparative studies among the SP6, T7, T3, and P4 proteins will also aid in the elucidation of the biochemical mechanism of primases and advance functional genomics. A fifth protein, gp37 from the lamdoid phage N15 (accession no. AAC48876), has recently been reported that also contains significant homology to both helicases and primases. If gp37 does possess primase or helicase activities, a biochemical analysis could further help define essential primase and helicase motifs and thus allow the more precise prediction of function from amino acid sequence.

[5] D. N. Frick, T. Y. Tseng, and C. C. Richardson, unpublished data.

ACKNOWLEDGMENT

We thank Benjamin B. Beauchamp, James M. Stattel, and Stanley Tabor for providing purified proteins and helpful discussions.

REFERENCES

1. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., W. H. Freeman, New York.
2. Griep, M. A. (1995) *Indian J. Biochem. Biophys. 32,* 171−178.
3. Tabor, S., and Richardson, C. C. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78,* 205−209.
4. Mendelman, L. V., and Richardson, C. C. (1991) *J. Biol. Chem. 266,* 23240−23250.
5. Scherzinger, E., Lanka, E., and Hillenbrand, G. (1977) *Nucleic Acids Res. 4,* 4151−4163.
6. Romano, L. J., and Richardson, C. C. (1979) *J. Biol. Chem. 254,* 10476−10482.
7. Mendelman, L. V., Kuimelis, R. G., McLaughlin, L. W., and Richardson, C. C. (1995) *Biochemistry 34,* 10187−10193.
8. Yoda, K., and Okazaki, T. (1991) *Mol. Gen. Genet. 227,* 1−8.
9. Cha, T. A., and Alberts, B. M. (1986) *J. Biol. Chem. 261,* 7001−7010.
10. Bujalowski, W., Klonowska, M. M., and Jezewska, M. J. (1994) *J. Biol. Chem. 269,* 31350−31358.
11. LeBowitz, J. H., and McMacken, R. (1986) *J. Biol. Chem. 261,* 4738−4748.
12. Arai, K., and Kornberg, A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76,* 4308−4312.
13. Lu, Y. B., Ratnakar, P. V., Mohanty, B. K., and Bastia, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93,* 12902−12907.
14. Liu, C. C., and Alberts, B. M. (1981) *J. Biol. Chem. 256,* 2821−2829.
15. Hinton, D. M., and Nossal, N. G. (1987) *J. Biol. Chem. 262,* 10873−10878.
16. Dong, F., and von Hippel, P. H. (1996) *J. Biol. Chem. 271,* 19625−19631.
17. Frick, D. N., Baradaran, K., and Richardson, C. C. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95,* 7957−7962.
18. Richardson, C. C. (1983) *Cell 33,* 315−317.
19. Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol. 166,* 477−535.
20. Matson, S. W., Tabor, S., and Richardson, C. C. (1983) *J. Biol. Chem. 258,* 14017−14024.
21. Bernstein, J. A., and Richardson, C. C. (1989) *J. Biol. Chem. 264,* 13066−13073.
22. Bernstein, J. A., and Richardson, C. C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85,* 396−400.
23. Hine, A. V., and Richardson, C. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91,* 12327−12331.

24. . Kusakabe, T., Hine, A. V., Hyberts, S. G., and Richardson, C. C. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96,* 4295−4300.
25. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 10638−10642.
26. Ilyina, T. V., Gorabalenya, A. E., and Koonin, E. V. (1992) *J. Mol. Evol. 34,* 351−357.
27. Ziegelin, G., Scherzinger, E., Lurz, R., and Lanka, E. (1993) *EMBO J. 12,* 3703−3708.
28. Notarnicola, S. M., Park, K., Griffith, J. D., and Richardson, C. C. (1995) *J. Biol. Chem. 270,* 20215−20224.
29. Butler, E. T., and Chamberlin, M. J. (1982) *J. Biol. Chem. 257,* 5772−5778.
30. Kassavetis, G. A., Butler, E. T., Roulland, D., and Chamberlin, M. J. (1982) *J. Biol. Chem. 257,* 5779−5788.
31. Calendar, R. (1988) The Bacteriophages in *The Viruses* (Fraenkel-Conrat, H., and Wagner, R. R., Eds.) Vol. 1, Plenum Press, New York.
32. Patel, S. S., Hingorani, M. M., and Ng, W. M. (1994) *Biochemistry 33,* 7857−7868.
33. Kusakabe, T., and Richardson, C. C. (1997) *J. Biol. Chem. 272,* 12446−12453.
34. Matson, S. W., and Richardson, C. C. (1983) *J. Biol. Chem. 258,* 14009−14016.
35. Ames, B. N., and Dubin, D. T. (1960) *J. Biol. Chem. 235,* 769−775.
36. Tabor, S., Huber, H. E., and Richardson, C. C. (1987) *J. Biol. Chem. 262,* 16212−16223.
37. Kusakabe, T., and Richardson, C. C. (1997) *J. Biol. Chem. 272,* 5943−5951.
38. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1993) *J. Biol. Chem. 268,* 27208−27213.
39. Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82,* 1074−1078.
40. Jin, D. J., and Gross, C. A. (1988) *J. Mol. Biol. 202,* 45−58.
41. Yuzhakov, A., Kelman, Z., and O'Donnell, M. (1999) *Cell 96*, 153−163.
42. Sun, W., and Godson, G. N. (1998) *J. Mol. Biol. 276,* 689−703.
43. Nakai, H., and Richardson, C. C. (1988) *J. Biol. Chem. 263,* 9831−9839.
44. Mohanty, U., Searls, T., and McLaughlin, L. W. (1998) *J. Am. Chem. Soc. 120,* 8275−8276.
45. Hiasa, H., Sakai, H., Tanaka, K., Honda, Y., Komano, T., and Godson, G. N. (1989) *Gene 84,* 9−16.
46. Swart, J. R., and Griep, M. A. (1993) *J. Biol. Chem. 268,* 12970−12976.
47. Frick, D. N., and Richardson, C. C. (1999) *J. Biol. Chem. 273* (in press).
48. Romano, L. J., and Richardson, C. C. (1979) *J. Biol. Chem. 254,* 10483−10489.
49. Swart, J. R., and Griep, M. A. (1995) *Biochemistry 34,* 16097−16106.
50. Sheaff, R. J., and Kuchta, R. D. (1993) *Biochemistry 32,* 3027−3037.
51. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25,* 3389−3402.
52. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22,* 4673−4680.
53. Versalovic, J., and Lupski, J. R. (1993) *Gene 136,* 281−286.
54. Aravind, L., Leipe, D. D., and Koonin, E. V. (1998) *Nucleic Acids Res. 26,* 4205−4213.
55. Mendelman, L. V., Beauchamp, B. B., and Richardson, C. C. (1994) *EMBO J. 13,* 3909−3916.
56. Kusakabe, T., and Richardson, C. (1996) *J. Biol. Chem. 271,* 19563−19570.
57. Notarnicola, S. M., and Richardson, C. C. (1993) *J. Biol. Chem. 268,* 27198−21207.
58. Patel, S. S., and Hingorani, M. M. (1993) *J. Biol. Chem. 268,* 10668−10675.
59. Guo, S., Tabor, S., and Richardson, C. C. (1999) *J. Biol. Chem. 274,* 30303−30309.

BI992155T