Essential Lysine Residues in the RNA Polymerase Domain of the Gene 4 Primase-Helicase of Bacteriophage T7*

Received for publication, August 31, 2001, and in revised form, October 17, 2001
Published, JBC Papers in Press, October 22, 2001, DOI 10.1074/jbc.M108443200

Seung-Joo Lee and Charles C. Richardson‡

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

At a replication fork DNA primase synthesizes oligoribonucleotides that serve as primers for the lagging strand DNA polymerase. In the bacteriophage T7 replication system, DNA primase is encoded by gene 4 of the phage. The 63-kDa gene 4 protein is composed of two major domains, a helicase domain and a primase domain located in the C- and N-terminal halves of the protein, respectively. T7 DNA primase recognizes the sequence 5′-NNGTC-3′ via a zinc motif and catalyzes the template-directed synthesis of tetraribonucleotides pppACCC, pppACCA, and pppACAC at primase recognition sites. Gene 4 primase synthesizes functional tetraribonucleotides that serve as primers for T7 DNA polymerase (11, 12). Gene 4 primase synthesizes functional tetraribonucleotide primers pppACCC, pppACCA, and pppACAC at primase recognition sequences 5′-GGGTC-3′, 5′-TGTTGTC-3′, and 5′-GTGTC-3′, respectively (13, 14). These primase recognition sites contain the basic recognition sequence 5′-GTC-3′ in which the 3′-cytidine is required for recognition but is not copied into the primer (15). At the sequence 5′-GTC-3′ gene 4 protein catalyzes the synthesis of diribonucleotide, pppAC, but only the tetraribonucleotides function as primers for T7 DNA polymerase (11, 16).

For concomitant progression with the leading strand during DNA replication, the lagging strand is synthesized discontinuously in the form of Okazaki fragments (1). For the recurrent initiation of Okazaki fragment synthesis in most prokaryotes, the requisite RNA primers are provided by DNA primase activity at specific DNA sequences on the single-stranded DNA (ssDNA) template generated by DNA helicase activity during leading strand DNA synthesis (2). In the bacteriophage T7 replication system, RNA primers are synthesized by the multifunctional protein encoded by gene 4 of the phage. The gene 4 protein provides two essential functions at the T7 replication fork: helicase and primase activities (2).

The helicase and primase domains of gene 4 protein are located in the C- and N-terminal halves, respectively, of the 566-amino acid polypeptide (see Fig. 1). A 26-amino acid segment linking the two domains is essential for the oligomerization of the protein into its functional state as a hexamer (3–7). The helicase domain enables the protein to translocate 5′ to 3′ on ssDNA by using the energy of dTTP hydrolysis (8). Upon encountering duplex DNA, the gene 4 protein unwinds the DNA strands progressively (9, 10).

The N-terminal half of gene 4 protein functions as a primase to catalyze the synthesis of template-directed oligoribonucleotides that serve as primers for T7 DNA polymerase (11, 12). Gene 4 primase synthesizes functional tetraribonucleotide primers pppACCC, pppACCA, and pppACAC at primase recognition sequences 5′-GGGTC-3′, 5′-TGTTGTC-3′, and 5′-GTGTC-3′, respectively (13, 14). These primase recognition sites contain the basic recognition sequence 5′-GTC-3′ in which the 3′-cytidine is required for recognition but is not copied into the primer (15). At the sequence 5′-GTC-3′ gene 4 protein catalyzes the synthesis of diribonucleotide, pppAC, but only the tetraribonucleotides function as primers for T7 DNA polymerase (11, 16).

Both the primase (17) and helicase (5, 18) domains have been purified from cells harboring cloned fragments of the respective coding regions of gene 4. Polypeptides derived from the C-terminal half of the protein form hexamers and exhibit helicase activity when those fragments contain the linker region (5, 18). Crystal structures of the helicase domain have shown that the T7 helicase is a member of the RecA family of helicases (6). The N-terminal primase fragment synthesizes oligoribonucleotides as efficiently as the full-length protein does in the absence of dTTP. However, the primase activity of the full-length gene 4 protein is stimulated by dTTP, a consequence of the helicase domain to which it is tethered (17). Because the primase domain binds weakly to ssDNA, it can be tethered to DNA via the hexameric helicase domain, resulting in a stimulation of primer synthesis (19). In addition, on large DNA molecules the translocation activity of the helicase serves to transport the primase domain to primase recognition sites (20). Finally, at least a portion of the helicase domain is required for interaction of the protein with T7 DNA polymerase for extension of the tetraribonucleotide primer by T7 DNA polymerase (17, 21).

Although a three-dimensional structure is not yet available for the T7 DNA primase domain of gene 4 protein, crystal structures of portions of other prokaryotic DNA primases in the same family have been obtained (22–24). A comparison of the amino acid sequence of gene 4 protein with DNA primases from bacteria and bacteriophages has revealed considerable homology within the primase family (25). The sequence alignment suggests there are six conserved motifs, indicated as motifs I through VI in Fig. 1 (see below). The most striking homology is found in motif I, which contains the Cys4 zinc motif. The crystal structure of the zinc motif of the primase of Bacillus stearo-
thermophilus has shown that it is a member of the zinc ribbon subfamily of zinc binding motifs such as that found in the yeast RNA polymerase II subunit 9 (22). Gene 4 protein contains 1 g atom of zinc per mole of gene 4 protein (26). *In vitro* mutagenesis of residues in this region as well as studies with chimeric T7 primases have demonstrated a role of the zinc motif in recognition of the trinucleotide sequence 5′-GTC-3′ (27, 28). Interestingly, a truncated species of gene 4 protein lacking the zinc motif is found in phage-infected cells in approximately equal molar amounts with the full-length gene 4 protein (29). The two co-linear proteins are translated from separate in-frame translational start sites (29). The larger 63-kDa protein depicted in Fig. 1 has both helicase and primase activities (30). The smaller 56-kDa protein lacks the N-terminal 63-amino acid residues found in the full-length protein. The 56-kDa protein is unable to catalyze template-directed oligoribonucleotide synthesis but has full helicase activity (31, 32). Both species of gene 4 proteins form hexamers in the presence of dTTP via interactions of the subunits with the linker region connecting the helicase and primase domains (5, 7, 33, 34). The 56-kDa gene 4 protein can catalyze the synthesis of random diribonucleotides at a very low rate (31). This observation also places the catalytic site for phosphodiester bond synthesis in the interior region of the primase domain.

In addition to the zinc motif that is conserved among prokaryotic primases, several other conserved sequence motifs are clustered in the central region of the primase domain (motifs II through VI). No role has been assigned to motif II but several residues within the remaining motifs appear to be critical for NTP binding and phosphodiester bond formation. Overlapping motif III is a charged, basic region found in several large subunits of both prokaryotic and eukaryotic RNA polymerases (35). Motifs IV, V, and VI have been implicated in the coordination of the NTP substrates and divalent metal cations necessary for catalysis (36–39). The x-ray crystal structure of the RNA polymerase domain of the Escherichia coli primase, containing motifs II through VI, shows that the residues that are strictly conserved in all bacterial DnaG proteins cluster around the central crevice of the protein (23, 24). One of these residues, Lys-241, plays an essential role in catalysis (40) and is one of the residues that can be cross-linked to NTP analogs (41). The center region of the protein contains a TOPRIM fold found in topoisomerases (42).

To date, only the effect of alterations in the zinc motif of gene 4 protein on primase activity has been examined. As discussed above, elimination of the zinc motif, for instance, as in the 56-kDa gene 4 protein, abolishes DNA-dependent primase activity but not the ability to synthesize random diribonucleotides (31). Less drastic changes, such as the substitution of any of the four cysteines of the zinc motif with serines, are lethal. One of the altered primases has been shown to lack primase activity (26). Mutations in the loop structure of the zinc motif and the adjacent region affect the recognition of the trinucleotide recognition sequence (27, 28). In the present study we have identified two essential lysines near motif III in the basic region of the RNA polymerase subdomain of gene 4 protein. The purified altered primases are defective in phosphodiester bond synthesis but not in their ability to identify primase recognition sites and to transfer a functional primer to T7 DNA polymerase.

**EXPERIMENTAL PROCEDURES**

**Materials**—Oligonucleotides were obtained from the Biopolymer Laboratory at Harvard Medical School. Restriction endonucleases, alkaline phosphatase, Deep Vent polymerase, and M13mp18 ssDNA were purchased from New England BioLabs. T7 polynucleotide kinase, T4 DNA ligase, radiolabeled nucleotides, and high molecular weight protein size markers were purchased from Amersham Biosciences, Inc. Agarose and β,γ-methylene dTTP were from United States Biochemical Corp. Polyethyleneimine cellulose thin layer chromatography (TLC) plates were from J. T. Baker. *E. coli* strain DH5α was from Invitrogen and HMS 174(DE3) was from Novagen. T7 DNA polymerase (T7 gene 5 protein-E. coli thioredoxin complex) was kindly provided by Donald Johnson (Harvard Medical School).

**Construction of Plasmids**—pET24gp4–63 was constructed by David Frick (Harvard Medical School) by inserting a gene 4-coding DNA fragment into pET 24a (Novagen) between *Hind*III and *Nde*I sites. In this gene 4-coding plasmid, the internal start codon at position 64 was replaced with the codon for glycine to avoid co-production of the 56-kDa gene 4 protein (43). The gene 4 protein-M64G contains all the catalytic properties of gene 4 protein encoded by wild-type T7 phage (44) and is referred as the 63-kDa wild-type gene 4 protein throughout this study. Mutations were introduced into this plasmid by overlap extension PCR in two steps (45, 46). The first round of PCR was performed with a pair of primers consisting of a mutagenic primer carrying a single mutation and an outside primer. Resulting PCR products were purified on a 1% agarose gel and were used as templates for the second round of PCR with a pair of outside primers. Products from the second PCR were digested with both *Apo*I and *Bst*I, and ligated with pET24gp4–63 previously cut with the same restriction enzymes. Mutated plasmids were transformed into *E. coli* DH5α, and the gene 4-coding regions were confirmed by DNA sequence analysis.

**Primase Overproduction and Purification**—Recombinant gene 4 proteins were purified following procedures described previously with the indicated modifications (47, 48). *E. coli* strain HMS 174(DE3), which contains the gene 4 protein-expressing plasmid, was grown to an *A*₅₄₀ of 1 in LB medium. Isopropyl β-D-thiogalactopyranoside was added to a final concentration of 1 mM. The cells were cultured for three additional hours, and then were harvested by centrifugation. Harvested cells were resuspended in buffer L (20 mM Tris–HCl, pH 7.5, 5 mM EDTA, 0.1 mM NaCl, 1 mM phenylmethylsulfonfluoride) and were subjected to three cycles of freezing and thawing in the presence of 0.2 mg/ml lysozyme. The lysed cells were centrifuged at 15,000 × *g* for 30 min, and polyethylene glycol (Fluka, PEG4000) was added to a final concentration of 10%. The PEG pellet was collected by centrifugation at 5,000 × *g* for 20 min, resuspended in buffer F (20 mM potassium phosphate, pH 6.8, 1 mM EDTA, 1 mM dithiothreitol (DTT), 10% glycerol), and loaded onto a phosphocellulose column (Whatman). Protein was eluted with a KC1 gradient from 0.2 to 1 M, and fractions containing gene 4 protein were combined. Fractions containing gene 4 protein were identified by gel analysis of an aliquot of each fraction. After the addition of MgCl₂ to a final concentration of 10 mM, the pooled fractions were loaded onto an Affi-Dexose affinity column (Sigma Chemical Co.) and eluted with buffer AE (20 mM potassium phosphate, pH 6.8, 20 mM EDTA, 0.5 mM DTT, 10% glycerol, 0.5 mM KC1).

In some cases, altered gene 4 proteins were purified using DEAE-Sephalac chromatography (Amersham Biosciences, Inc.). Pooled protein from the phosphocellulose column was dialyzed against buffer D (20 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 0.5 mM DTT, 10% glycerol) and loaded onto a phosphocellulose column (Whatman). Protein was eluted with a KC1 gradient from 0.2 to 0.7 M, and the fractions containing gene 4 protein were combined. Fractions containing gene 4 protein were identified by gel analysis of an aliquot of each fraction. After the addition of MgCl₂ to a final concentration of 10 mM, the pooled fractions were loaded onto an Affi-Dexose affinity column (Sigma Chemical Co.) and eluted with buffer AF (20 mM potassium phosphate, pH 7.5, 0.1 mM EDTA, 0.1 mM DTT, 50% glycerol) and stored at −20 °C until use.

**Primase Oligoribonucleotide Synthesis Assay**—The *de novo* synthesis of oligoribonucleotides catalyzed by the gene 4 primase was determined by measuring the incorporation of radioactively labeled CTP into oligoribonucleotides using a synthetic DNA template containing a primase recognition site (15, 19). The reaction (10 μl) included the indicated amount of template (5′-GGTTCA-3′), 0.1 μM each of ATP and CTP, and 0.1 μCi of [α-32P]CTP, 50 mM (monomeric concentration) gene 4 protein, 40 mM Tris–HCl, pH 7.5, 10 mM MgCl₂, 10 mM DTT, and 50 mM potassium glutamate. After incubation at 37 °C for 20 min, the reaction was terminated by the addition of 3 μl of sequencing dye (98% formamide, 10 mM EDTA, pH 8.0, 0.1% xylene cyanol FF, and 0.1% bromphenol blue) and loaded onto a 25% denaturing polyacrylamide sequencing gel containing 3 μl urea. Radioactive oligoribonucleotide products were analyzed using a Fuji BAS 1000 Bioimaging analyzer.
units of alkaline phosphatase for 20 min. The reaction was terminated by the addition of 3 μl of sequencing dye, and the products were separated on a 25% denaturing polyacrylamide sequencing gel containing 3 μM urea.

**Primase Oligoribonucleotide Extensions Assay**—Gene 4 protein can catalyze the extension of the 5'-rAC-3' or 5'-rACC-3' at primase recognition sites on a template in the presence of the appropriate NTPs (16). The reaction mixture was as described above for the primase oligoribonucleotide synthesis assay except that the indicated concentration of 5'-rACC-3' or 5'-rACC-3' was used. The reaction was carried out and analyzed as described in the assay for primase oligoribonucleotide synthesis.

**RNA-primer DNA Synthesis Assay**—The ability of gene 4 protein to prime DNA synthesis catalyzed by T7 DNA polymerase on M13 ssDNA was measured as previously described (10, 16, 49). The reaction (10 μl) contained 9.8 nM M13 ssDNA (molecular concentration of DNA), 0.3 mM of all four dNTPs, 0.1 mM of ATP and CTP, or supplied by the addition of 0.01 mM of the tetraribonucleotide rACCC. After incubation for 10 min at 37 °C, the reaction was terminated by the addition of EDTA to a final concentration of 25 mM. The reaction mixture was loaded onto a 10% non-denaturing gel, and electrophoresed at 4 °C for 20 min, the reactions were analyzed using a Fuji BAS 1000 Bioimaging analyzer.

**dTTPase Assay**—Gene 4 protein catalyzes the ssDNA-dependent hydrolysis of dTTP, a reaction coupled to its translocation on ssDNA (8). The reaction mixture was as described above for the primase oligoribonucleotide synthesis assay except that the indicated concentration of 5'-rACC-3' or 5'-rACC-3' was used. The reaction was carried out and analyzed as described in the assay for primase oligoribonucleotide synthesis.

**DNA Binding Assay**—DNA binding affinity of gene 4 protein was measured by nitrocellulose (NC) filter binding. The reaction (10 μl) containing 1 nM 5'-end-radio-labeled DNA (5'-CAAGGCACGCGTAG-3'), 40 mM Tris-HCl, pH 7.5, 10 mM Mg(OAc)2, 10 mM DTT, and 50 mM potassium glutamate. After incubation for 20 min at 37 °C, the reaction mixture was loaded onto a 10% non-denaturing gel, and electrophoresed at 4 °C for 20 min, the reactions were analyzed using a Fuji BAS 1000 Bioimaging analyzer.

**RESULTS**

The “RNAP-basic” motif derived from an alignment of bacterial primases with the large subunits of DNA-dependent RNA polymerase is likely to play an important role in the RNA synthesis catalyzed by DNA primases (35). Although the amino acid sequence of T7 gene 4 protein does not align directly with the RNAP-basic region, the gene 4 protein does have basic and hydrophobic residues in this region. Specifically, there are six basic residues located within the relatively short segment from amino acid position 122 to 137, immediately adjacent to motif III (Fig. 1). In an attempt to identify residues that contribute catalytic function to the gene 4 primase, we have replaced five lysines in this region one by one with alanines and have then examined the effect of these alterations on the properties of the gene 4 protein.

**Complementation Analysis**—We examined the ability of the altered gene 4 proteins to complement the function of gene 4 protein in vivo using a phage complementation assay. T7Δ4–1, a phage in which the gene 4-coding region has been deleted, can grow only in E. coli cells harboring a plasmid that expresses a functional gene 4 (43). Plasmids expressing the five altered gene 4 proteins were constructed and transformed into E. coli DH5α. The host cells were then infected with T7Δ4–1. The results of the assay demonstrate that the gene 4 protein in which alanine was substituted for lysine 122 (gp4-K122A) does not support the growth of the gene 4-deleted T7 phage (Table I). Gene 4 proteins with alterations of K126A and K128A (gp4-K126A and gp4-K128A) exhibit diminished ability to complement the phage. Two other gene 4 proteins, gp4-K131A and gp4-K137A, complement T7Δ4–1 as efficiently as does the wild-type gene 4 protein (Table I). Thus, alteration of gene 4 protein at either Lys-122, Lys-126, or Lys-128 disturbs an essential function of gene 4 protein in vivo. However, alteration of gene 4 protein at either Lys-122, Lys-126, or Lys-128 disturbs an essential function of gene 4 protein in vivo. However, alteration of gene 4 protein at either Lys-122, Lys-126, or Lys-128 disturbs an essential function of gene 4 protein in vivo.
of two other lysines in this region, Lys-131 or Lys-137, does not impair gene 4 protein function.

Some altered gene 4 proteins have been shown to be dominantly lethal for wild-type T7 phage growth (5, 47, 50). To examine if alteration of lysines in this region causes dominant lethality, the same host E. coli cells expressing the altered gene 4, as described above, were infected with wild-type T7 phage. No reduction in viability of the phage was observed (Table I), indicating that the altered gene 4 proteins do not interfere with gene 4 protein produced by the wild-type phage.

**Overproduction and Purification of the Altered Gene 4 Proteins**—To analyze the DNA primases biochemically, each of the five altered proteins was overproduced from cells harboring their cloned genes, and the proteins were then purified according to established protocols (47, 48). The wild-type and altered proteins were overproduced to the same extent in E. coli HMS 174(DE3) containing the cloned gene 4 and behaved similarly in the purification. As judged by SDS gel analysis, the purity of all of the altered proteins was greater than 95%. Overall, purification yields of all of the altered gene 4 proteins were similar to that obtained with the wild-type protein. To rule out any differences in enzyme activity that could be attributed to the alternate purification procedure, we separately purified the wild-type gene 4 protein and gp4-K128A using both an ATP-agarose and a DEAE-cellulose column. As judged by activity in a standard primase oligoribonucleotide synthesis assay, no disparity between the proteins purified by either method was found.

**Oligoribonucleotide Synthesis**—The results obtained from complementation analysis suggested that alteration of specific lysine residues within the primase domain disrupts a vital function of T7 DNA primase. To determine the biochemical basis of these defects, we first measured the ability of each of the altered proteins to catalyze template DNA-directed oligoribonucleotide synthesis. In this assay the ability of the primase to catalyze the synthesis of di-, tri-, and tetraribonucleotide from ATP and CTP at specific recognition sequences in a ssDNA template was measured (15). In the assay shown in Fig. 2, the 7-nucleotide template contained the primase recognition site 5′-GGGTG-3′. On this template, in the presence of ATP and [α-32P]CTP, gene 4 protein catalyzed the synthesis of pppAC, pppACC, and pppACCC, a reaction that is proportional to the amount of template. This activity was determined by analysis of the radioactive ribonucleotides on a denaturing polyacrylamide gel (Fig. 2). A small amount of pentaribonucleotide, pppACCCC and pppACCCCA, was also observed, presumably arising from misincorporation of the fifth ribonucleotide (15, 19). In striking contrast, no oligoribonucleotide synthesis was observed with gp4-K122A even at the highest concentration of template (Fig. 2). The other four altered gene 4 proteins all synthesized oligoribonucleotides, although the amount of synthesis obtained with gp4-K128A, gp4-K131A, and gp4-K137A appears to be decreased relative to the wild-type gene 4 protein and gp4-K126A.

To obtain more quantitative data on the amount of synthesis catalyzed by each of the proteins other than gp4-K122A, we measured the amount of radioactive oligoribonucleotides in each lane of the gel (Fig. 2). The results confirm the visual impression that gp4-K126A appears to be identical to the wild-type gene 4 protein, whereas the other three proteins show significantly reduced synthesis. When M13mp18 ssDNA, which contains a total of 68 sites of 5′-GTC-3′ sequences, was used as the template, similar results in primer synthesis were obtained.
observed (data not shown). Analysis of the oligoribonucleotide products synthesized on M13 ssDNA indicates that all the altered proteins except for gp4-K122A recognize the same sites in the template as does the wild-type protein.

**DNA Template-independent Diribonucleotide Synthesis**

Earlier studies on the gene 4 protein revealed that the 56-kDa gene 4 protein, lacking the zinc motif, catalyzed the random synthesis of diribonucleotide in the presence of all four NTPs (31). Subsequent studies have shown that this reaction occurs in the absence of a DNA template and yields diribonucleotides having essentially any ribonucleotide at the 5'-end and a preference for cytosine at the 3'-end (28). It has been postulated that this activity is generated by the catalytic site of gene 4 primase that binds and condenses NTPs (26, 51). DNA-independent synthesis using the wild-type gene 4 protein is shown in Fig. 3 where diribonucleotide products are apparent. Among the five altered gene 4 proteins, only gp4-K122A and gp4-K128A were defective in the synthesis of diribonucleotides (Fig. 3). The other three genetically altered proteins were indistinguishable from the wild-type gene 4 protein in their activity to catalyze the DNA-independent synthesis of diribonucleotides.

**Oligoribonucleotide Extensions**

In the oligoribonucleotide synthesis assay, the gene 4 protein catalyzes the de novo synthesis of tetraribonucleotides from the precursor ATP and CTP at primase recognition sites. Gene 4 protein can also extend a diribonucleotide, rAC, and a triribonucleotide, rACC, to the tetraribonucleotide rACCC in the presence of CTP and a template containing the recognition sequence 5'-GGGTC-3' (16). The 3'-cryptic cytosine is essential for the extension reaction.

We examined the ability of the genetically modified gene 4 proteins to mediate this extension reaction. The extension activity was measured by incubating the gene 4 protein with either the diribonucleotide rAC or the triribonucleotide rACC in the presence of CTP and a template containing the recognition sequence site 5'-GGGTC-3'. The reaction products were identified by denaturing polyacrylamide gel analysis, and the quantita-

---

![Fig. 3. DNA-independent diribonucleotide synthesis by gene 4 protein.](https://example.com/fig3)

![Fig. 4. Oligoribonucleotide extension by gene 4 protein at various oligoribonucleotide concentrations.](https://example.com/fig4)
tion data are presented in Fig. 4. The ability of the wild-type gene 4 protein to extend a diribonucleotide and a triribonucleotide is apparent in Fig. 4. Gp4-K122A, by contrast, was unable to extend either preformed oligoribonucleotide. Gp4-K128A had a very low but detectable activity, whereas the remaining three altered proteins showed an activity equivalent to, or higher than, the wild-type protein. These results extend those obtained with the DNA-independent diribonucleotide synthesis assay in that we can conclude that lysines 122 and 128 play an essential role not only in the formation of the first phosphodiester bond but in the additional condensations as well.

**RNA-primed DNA Synthesis**—The major function of the T7 DNA primase, like all DNA primases, is to provide oligoribonucleotides as primers for DNA polymerase to initiate DNA synthesis. T7 DNA polymerase uses the tetraribonucleotides synthesized by the gene 4 protein as primers to initiate DNA synthesis (11). T7 DNA polymerase itself does not efficiently use a tetraribonucleotide as a primer (16, 52). A preformed tetraribonucleotide supplied exogenously can also be utilized by the gene 4 protein to prime DNA synthesis at a primase recognition site (16, 53).

In the experiment shown in Fig. 5, we have assayed RNA-primed DNA synthesis catalyzed by T7 DNA polymerase under different assay conditions. These conditions are depicted schematically in Fig. 5A. First, gene 4 protein was examined for its activity to synthesize functional primers de novo from ATP and CTP on a template containing the recognition sequence 5′-GGGTC-3′. As shown in Fig. 5B (open bars), the wild-type gene 4 protein as well as the altered proteins, gp4-K126A, gp4-K131A, and gp4-K137A all synthesized functional primers. In contrast, gp4-K122A and gp4-K128A did not.

In the second assay, the preformed diribonucleotide rAC and CTP were provided along with the same template (Fig. 5A). Although this reaction was less efficient (solid bars), the pattern was similar to that observed with ATP and CTP. Finally, in the third assay, a synthetic tetraribonucleotide rACCC was provided to the gene 4 protein in the absence of ATP and CTP (Fig. 5A). All of the gene 4 proteins, including gp4-K122A and gp4-K128A, were able to transfer the primer to T7 DNA polymerase (hatched bars). We conclude that gp4-K122A and gp4-K128A are able to recognize both the tetraribonucleotide and the primase recognition site. Furthermore, these altered proteins can properly interact with T7 DNA polymerase and transfer a functional primer. Therefore, the basic defect appears to be in synthesis of phosphodiester bonds.

**Properties of the Helicase Domain in the Altered Gene 4 Proteins**—The amino acid substitutions made in the primase domain of gene 4 protein are not expected to influence the helicase activity of the C-terminal half of the protein. We have examined several properties attributed to the helicase domain: dTTP hydrolysis, DNA binding, unwinding of duplex DNA, and hexamer formation (Table II). ssDNA-dependent hydrolysis of dTTP is indicative of the unidirectional translocation of gene 4 protein on ssDNA (8). Because translocation is in turn dependent on the binding of the gene 4 protein to ssDNA and on hexamer formation, dTTPase activity provides a sensitive measure of overall helicase functions. As shown in Fig. 6, all five genetically altered gene 4 proteins displayed ssDNA-dependent dTTPase activity equivalent to the wild-type gene 4 protein. In confirmation of the dTTPase assay, no notable difference between the altered proteins and the wild-type gene 4 protein was found with regard to binding to ssDNA, unwinding of duplex DNA, and oligomerization (Table II). These results confirm our interpretation that the single amino acid changes in the primase domain do not affect helicase activity.

**DISCUSSION**

The bacterial and phage DNA primases, although distantly related, share many structural and biochemical properties, including a close association with DNA helicases and numerous signature sequences (2, 25, 35, 42). The three extensively studied prokaryotic primases, the DnaG protein of *E. coli*, the gene 61 protein of phage T4, and the gene 4 helicase/primase of phage T7, illustrate this point. All three DNA primases have a zinc motif in which the metal is coordinated by four cysteines in the T4 and T7 primases and by three cysteines and a histidine in the *E. coli* DnaG primase (25, 26, 54). All three primases recognize a specific trinucleotide sequence in ssDNA, a sequence that differs for each primase, with the 3′-nucleotide of the sequence being essential for recognition but not copied into the product oligoribonucleotide (15, 55, 56). In the case of the T7 primase the zinc motif has been shown to play an important role in the recognition of the basic trinucleotide sequence (26–
Results were obtained from duplicated assays carried out as described under "Experimental Procedures." Kinetic constants were derived using the Macintosh computer program Enzyme Kinetics (Trinity Software).

|                 | WT     | K122A | K126A | K128A | K131A | K137A |
|-----------------|--------|-------|-------|-------|-------|-------|
| dTTP hydrolysis |
| $K_a$ (mM)      | 3.3 ± 0.3 | 1.8 ± 0.3 | 2.2 ± 0.1 | 4.2 ± 0.2 | 2.8 ± 0.5 | 1.3 ± 0.1 |
| $V_{max}$ (pmol of dTTP hydrolyzed/second) | 1.3 ± 0.04 | 1.2 ± 0.02 | 1.4 ± 0.1 | 1.7 ± 0.1 | 1.2 ± 0.1 | 1.1 ± 0.1 |
| DNA binding     |
| $K_d$ (nm)      | 25 ± 2  | 19 ± 3 | 19 ± 2 | 38 ± 5 | 28 ± 4 | 27 ± 4 |
| DNA unwinding   |
| ssDNA (pmol)    | 0.2 ± 0.08 | 0.2 ± 0.01 | 0.3 ± 0.07 | 0.2 ± 0.07 | 0.2 ± 0.06 | 0.3 ± 0.05 |
| Hexamer formation | +   | +   | +   | +   | +   | +   |

- $K_a$ and $V_{max}$ were derived from data as shown in Fig. 6.
- $T_{dissoc}$ association constant for binding of gene 4 protein to DNA.
- Amount of unwound DNA by 0.01 pmol of gene 4 protein from initial 2 pmol of DNA substrate.

FIG. 6. dTTP hydrolysis by gene 4 protein. Gene 4 protein catalyzes the ssDNA-dependent hydrolysis of dTTP to dTDP and P$_i$. The reaction contained 1.1 nM M13 ssDNA, the indicated concentration of dTTP (0.5, 1, 2, 3, 4, 6, 8, 10 mM), 0.1 μCi of [$\gamma$-32P]dTTP, and 80 nM of the indicated gene 4 protein (see "Experimental Procedure"). After incubation at 37 °C for 20 min, the products of the reaction were analyzed by TLC as described under "Experimental Procedure." The rate of dTTP hydrolysis is plotted against the initial concentration of dTTP. WT (●), K122A (△), K126A (♦), K128A (□), K131A (○), K137A (○).

The final functional primers synthesized by the primases have lengths dependent on the specific primase: tetranucleotides, pentanucleotides, and predominantly lengths of eleven nucleotides for the T7, T4, and E. coli primases, respectively (57–60). Essential interactions with DNA helicase at the replication fork have been documented with all three proteins but the T7 primase is unique among the three in that it is a part of the same polypeptide that contains the DNA helicase domain.

Except for RNAP-basic motif (see Fig. 1), DNA primases do not have significant homology with DNA-dependent RNA polymerases. Rather, primases have been predicted to have structural similarity to the more functionally distant topoisomerases comprising the TOPRIM superfamily (42). A recent crystal structure of the catalytic core of the E. coli DnaG primase revealed that alternating α helices and β sheets in the central part of the protein create an acidic metal binding site with a fold similar to that found in topoisomerases (23, 24). Consisting of three β strands, the region around RNAP-basic motif creates a depression of the basic surface adjacent to the metal binding site. Interestingly, invariant residues in catalytic domain of bacterial primases cluster at a concave region generated by the TOPRIM and the neighboring basic region. The recognition site for the E. coli DnaG protein on ssDNA is 5'-CTC-3' with initial synthesis yielding pppAG (55). Affinity labeling of the catalytic site with ATP analogs identified three lysine residues at positions 211, 229, and 241, all of which are located near RNAP-

basic motif (41). However, only at position 241, does replacement of lysine with arginine result in an altered primase that aborts synthesis after the initial diribonucleotide product; substitution at position either 211 or 229 has no effect (40). Lysine 241 is located at the junction between the basic depression that includes the RNAP-basic motif and the metal binding site created by TOPRIM fold.

Our attempts to identify the catalytic site of the T7 DNA primase were based in part on these studies of the E. coli DnaG protein. Earlier alignment of the primase sequences of these two proteins had suggested that lysine 137 in the T7 gene 4 protein corresponded to the critical lysine 241 in the E. coli DnaG protein (25). However, in the present study we find that the lysines at positions 122 and 128 but not the one at 137 play a crucial role at the active site. Our own alignment of the DnaG and T7 gene 4 protein using the program ClustalX (61) is almost identical to that reported earlier (25) except for residues in motif III and the overlapping RNAP-basic region. Based on the results we report in this study, we extend the basic region in the T7 primase further toward the C terminus. This alignment places the essential lysines at positions 122 and 128 adjacent to the metal binding site as defined in the DnaG sequence alignment.

In the absence of a three-dimensional structure for the T7 primase domain, it is difficult to assign specific roles to lysines 122 and 128. Because the amino acid sequence of the gene 4 protein in the vicinity of these residues is also heavily basic, they may be part of an electropositive cleft similar to that found in the DnaG structure. It has been proposed that this basic cleft found in the E. coli DnaG primase interacts with the backbone of the ssDNA template (23, 24). However, it is hard to relate the inability of the altered gene 4 proteins to catalyze diribonucleotide synthesis in the absence of DNA to a defect in DNA binding. The DnaG affinity-labeling studies with ATP analogs mentioned above provided evidence that the 5'-terminal ATP of the oligonucleotide being synthesized remains bound at the active site (41). However, at least in the case of the T7 primase, it seems unlikely that a 5'-phosphate on the oligoribonucleotide is essential in view of the fact that the wild-type primase can bind and extend a preformed diribonucleotide lacking a 5'-phosphate (16), and can bind and transfer a preformed tetranucleotide lacking a 5'-phosphate to the DNA polymerase (this study).

On the other hand, the amino side chains of lysine are known to interact with the phosphate moiety of an incoming nucleotide in the active site of both DNA polymerases (62, 63) and RNA polymerases (64, 65). This is a distinct possibility for the lysines identified in the present study. Based on kinetic studies
carried out with the N-terminal fragment of gene 4 protein containing only the primase domain we have proposed that there are two NTP binding sites at the active site (51). In this model for the T7 primase the first nucleotide (ATP) binds to the initiation site, and the next incoming nucleotide (CTP) binds to the second or elongation site. At each elongation step the growing (n+1) primer must be transferred to the initiation site so that another NTP can bind to the elongation site. Most primases show little preference regarding the triphosphate end of the NTP incorporated at the 5′-end of the primer (2), and hence it is less likely that the essential lysines would be involved in interaction with the first nucleotide, ATP. If indeed their role is to interact with the 5′-triphosphate moiety, then it is likely that they play a role at our proposed elongation site where CTP is preferred.

The current study in combination with previous studies on the zinc motif of T7 gene 4 protein demonstrates a clear distinction between recognition of the primase recognition sequence and oligonucleotide synthesis. Earlier studies have shown that the 56-kDa gene 4 protein, lacking the zinc motif, is defective in DNA-dependent oligonucleotide synthesis but still shows that the 56-kDa gene 4 protein, lacking the zinc motif, is preferred.

Essential Lysine Residues in T7 DNA Primase

Acknowledgments—We are grateful to Joosoon Lee, David Frick, and Stan Tabor for helpful discussions and suggestion. We thank Lisa Rezende, Ingrid Richardson, and Tom Ellenberger for critical reading of the manuscript.

REFERENCES

1. Kornberg, A., and Baker, T. A. (1992) DNA Replication, 2nd Ed., W. H. Freeman, New York

2. Patel, S. S., and Picha, K. M. (2000) Annu. Rev. Biochem. 69, 651–697
3. Nakai, H., and Richardson, C. C. (1988) J. Biol. Chem. 263, 9818–9830
4. Romano, L. J., and Richardson, C. C. (1979) J. Biol. Chem. 254, 10438–10489
5. Romano, L. J., and Richardson, C. C. (1979) J. Biol. Chem. 254, 10478–10482
6. Tabor, S., and Richardson, C. C. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 205–209
7. Egelman, E. H., Yu, X., Wild, R., Hingorani, M. M., and Patel, S. S. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3869–3873
8. Patel, S. S., and Picha, K. M. (2000) Annu. Rev. Biochem. 69, 651–697
9. Nakai, H., and Richardson, C. C. (1988) J. Biol. Chem. 263, 9818–9830
10. Matson, S. W., and Richardson, C. C. (1983) J. Biol. Chem. 258, 14019–14016
11. Matson, S. W., Tabor, S., and Richardson, C. C. (1983) J. Biol. Chem. 258, 14017–14024
12. Godson, G. N., Schoenich, J., Sun, W., and Mustaev, A. A. (2000) Biochemistry 39, 322–339
13. Nakai, H., and Richardson, C. C. (1988) J. Biol. Chem. 263, 12469–12474
14. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
15. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
16. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
17. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
18. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
19. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
20. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
21. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
22. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
23. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
24. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
25. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
26. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
27. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
28. Mendelman, L. V., Notarnicola, S. M., and Richardson, C. C. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1228–1232
Essential Lysine Residues in the RNA Polymerase Domain of the Gene 4 Primase-Helicase of Bacteriophage T7
Seung-Joo Lee and Charles C. Richardson

J. Biol. Chem. 2001, 276:49419-49426.
doi: 10.1074/jbc.M108443200 originally published online October 22, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108443200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 64 references, 40 of which can be accessed free at http://www.jbc.org/content/276/52/49419.full.html#ref-list-1