Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 4295–4300, April 1999
Biochemistry

The Cys_4 zinc finger of bacteriophage T7 primase in sequence-specific single-stranded DNA recognition

TAKAHIRO KUSAKABE*†, ANNA V. HINE‡, SVEN G. HYBERTS*, AND CHARLES C. RICHARDSON*§

*Department of Biological Chemistry and Molecular Pharmacology, Harvard University Medical School, Boston, MA 02115; and ‡Pharmaceutical Sciences Research Institute, Aston University, Aston Triangle, Birmingham, B4 7ET, United Kingdom

Contributed by Charles C. Richardson, February 8, 1999

ABSTRACT **Bacteriophage T7 DNA primase recognizes 5′-GTC-3′ in single-stranded DNA. The primase contains a single Cys_4 zinc-binding motif that is essential for recognition. Biochemical and mutagenic analyses suggest that the Cys_4 motif contacts cytosine of 5′-GTC-3′ and may also contribute to thymine recognition. Residues His_{33} and Asp_{31} are critical for these interactions. Biochemical analysis also reveals that T7 primase selectively binds CTP in the absence of DNA. We propose that bound CTP selects the remaining base G, of 5′-GTC-3′, by base pairing. Our deduced mechanism for recognition of ssDNA by Cys_4 motifs bears little resemblance to the recognition of trinucleotides of double-stranded DNA by Cys_2His_2 zinc fingers.**

Single Cys_4 zinc-binding motifs participate in protein–nucleic acid interactions. Predominantly, they are found in enzymes involved in DNA replication/transcription. For example, bacteriophage T4 and T7 primases each contain a single Cys_4 motif and each recognize specific trinucleotides of single-stranded DNA (ssDNA) (1, 2), whereas peptide fragments containing the Cys_4 motif of transcription factor TFIIS interact with ssDNA (3, 4), double-stranded DNA and RNA (5). These motifs occur in other primases (2, 6, 7), transcription factor TFIIE (8), and in subunits of RNA polymerase II and bacterial DNA repair proteins (9). Structural analysis of the single Cys_4 zinc-binding motifs found in transcription factors TFIIS and TFIIB suggest that these "zinc ribbons" have little in common with other zinc-binding motifs: their secondary structure is almost entirely β-sheet (3, 9, 10).

T7 DNA primase has been characterized extensively and hence offers advantages for studying the Cys_4 motif. Primase is the product of T7 gene 4, a gene that encodes two colinear proteins, a 63-kDa helicase/primase (referred to simply as primase) and a 56-kDa helicase (11). Helicase is translated from an internal AUG codon and lacks the Cys_4 motif. It binds to ssDNA as a hexamer (12) and translocates 5′ to 3′ along the DNA by hydrolyzing dTTP (13, 14). On encountering a duplex region, the protein continues to translocate and thus unwinds the DNA.

The 63-kDa gene 4 primase/helicase possesses all of the above activities, but additionally catalyzes the template-directed synthesis of oligoribonucleotides on ssDNA. These are used as primers by T7 DNA polymerase (15–17). The functional primers pppACCC, pppACCA, and pppACAC are synthesized at the recognition sites 5′-GGGTC-3′, 5′-TGGTC-3′, and 5′-CTGTC-3′ (13). However, the minimal primase recognition site, 5′-GTC-3′, is sufficient for recognition and supports the synthesis of pppAC dimers (2). The 3′-cytidine is cryptic: it is required for recognition but is not copied into the primer.

There is homology between the 245 N-terminal amino acids of T7 primase and prokaryotic primases, whereas the C-terminal amino acids (residues 277–566) show homology to prokaryotic helicases (7). Recently, DNA sequences encoding these two domains have been cloned individually. The C-terminal domain has helicase activity (18) and contains the dTTPase site essential for binding to ssDNA and for translocation (19, 20) and the sites for hexamer formation and interaction with T7 DNA polymerase (21, 22). The N-terminal domain has only primase activity (23).

A cartoon depicting the 63-kDa gene 4 protein illustrates the dTTP binding site in the helicase domain and proposes the presence of two NTP binding sites in the primase domain, one for ATP and one for CTP (Fig. 1). Removal or disruption of the Cys_4 motif destroys recognition of 5′-GTC-3′ (6, 24). In contrast, a chimeric T7 primase containing the Cys_4 motif of bacteriophage T3 primase recognizes the same 5′-GTC-3′ site as native T7 primase. In this protein, the Cys_4 loop region is conserved, whereas surrounding sequences are highly substituted (25). However, chimeric T7 primases that contain only the Cys_4 loop region of either phage T4 primase or the *Escherichia coli* primase each recognize, albeit poorly, a novel trinucleotide sequence (26). Clearly, the Cys_4 motif is not the sole determinant of sequence specificity.

We have used mutational analysis and biochemical characterization of both native and altered primases to examine the role of the Cys_4 motif. The results suggest contacts between the Cys_4 motif and its trinucleotide recognition sequence. Our data imply that the 5′ and 3′ nucleotide binding sites of T7 primase contribute to sequence specificity.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. *E. coli* DH5α and BL21(DE3) and plasmid pGP4-6_{G64S10} have been described (27, 28). Plasmid pET4-6_{G64S10} was constructed by cloning the *Nde*I-*Bgl*II fragment of pGP4-6_{G64S10} into the *Nde*I-*Bam*HI sites of pET11b (Novagen). Each plasmid encodes a T7 primase in which Met_{64} is substituted with Gly. Primase $M_{64}G$ has all the catalytic activities of T7 primase (28), and hereafter is referred to as wild type. Bacteriophage T7Δ4–1 was provided by S. Tabor (Harvard Medical School) (29).

Enzymes and DNAs. All primases were purified from *E. coli* BL21(DE3) containing the expression vector pET4-6_{G64S10} (21). Enzymes were from Amersham Pharmacia or New England Biolabs. Sequences of oligonucleotide templates were as follows.

For analysis of the 3′ cryptic nucleotide: GT**N**30:15 (where **N** is alternatively A, C, G, and T): 5′-GGATTAGCGGAGA-

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

PNAS is available online at www.pnas.org.

Abbreviation: ssDNA, single-stranded DNA.

†Present address: Department of Agronomy, Faculty of Agriculture, Kyusyu University, Fukuoka 812-81, Japan.
§To whom reprint requests should be addressed at: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115. e-mail: ccr@bcmp.med.harvard.edu.

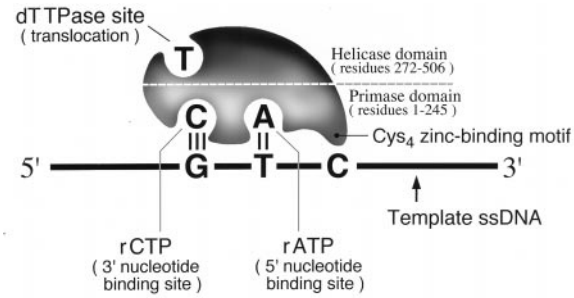

FIG. 1. Model of T7 primase/helicase. The helicase domain comprises residues 272–566 (7) and contains the dTTP binding site (19, 20). The primase domain, including the Cys_4 motif, comprises residues 1–245 (7) and recognizes the sequence 5′-GTC-3′ in ssDNA. Here it synthesizes the dinucleotide 5′-pppApC-3′ from ATP and CTP (23). We denote the ATP and CTP binding sites as the 5′ and 3′ nucleotide binding sites, respectively.

TTTGAGAGGATATGCGAGGGTNTTTTTTTTTTTTTT-TT-3′ and GTC^{x}5:10 (where C^{x} is alternatively C and ^{4H}C): 5′-GCTGATGGTC^{x}AGTGGTATCG-3′. For analysis of T analogs within 5′-GTC-3′: GT^{x}C5:10, (where T^{x} is alternatively T, dU, C, and ^{5m}C): 5′-CTATCGGGT^{x}CATCCCACAGG-3′.

Construction of Expression Vectors. Gene 4 mutants used solely for *in vivo* analyses of gene 4 were constructed from the vector pGP4-6_{G64S10} by the "2-primer" method (Strategene). Mutants used in both *in vivo* and *in vitro* analyses of gene 4 function were constructed in the plasmid pET4–6_{G64S10} by PCR. Full protocols for the construction of these mutants may be found in the supplemental material on the PNAS web site (www.pnas.org).

Enzyme Assays. Nucleotide hydrolysis assays were carried out as described (26) with 50 nM M13 DNA and 150 nm primase (monomer). Radioactivity was quantified by using a BAS100 Fuji Bio-Imaging analyzer. Oligonucleotide synthesis assays (30-min incubations) were carried out as described (30) with 50 nM primase (hexamer), 100 μM NTPs as indicated and 100 nM synthetic oligonucleotide template or 20 nM M13 DNA. In template-independent assays, the DNA was omitted, 10 mM $MnCl_2$ replaced the $MgCl_2$, and reactions were treated with calf intestinal alkaline phosphatase before electrophoresis (2).

RESULTS

Polymerization of NTPs by Primase. To synthesize the diribonucleotide pppAC, T7 primase must first bind ATP and CTP. Extension of a synthetic pppAC dinucleotide by primase suggested that the dinucleotide is bound before sequence recognition (30). Might the precursors, ATP and CTP, similarly be bound in the active site before template recognition? If so, the nucleotides might themselves play a role in ssDNA recognition. We examined nucleotide binding and polymerization in the absence of a ssDNA template. The 63-kDa primase catalyzes diribonucleotide synthesis, albeit at low levels, in the absence of a template (6), a reaction enhanced by manganese (24).

To examine the first phosphodiester bond formation, synthesis of dinucleotides was monitored in the absence of DNA with each of the four NTPs and $MnCl_2$. Because cytosine is preferentially incorporated into the second position in template-directed dinucleotide synthesis (pppAC), we monitored template-independent dinucleotide synthesis with [α-^{32}P]CTP. In the reaction, the 5′ triphosphate is removed by alkaline phosphatase, so that radiolabel remains only in the phosphodiester linkage. Hence, only dinucleotides that contain cytosine as the second base are detected. In the absence of template DNA, primase catalyzes the synthesis of AC (Fig. 2*A*, lane 1),

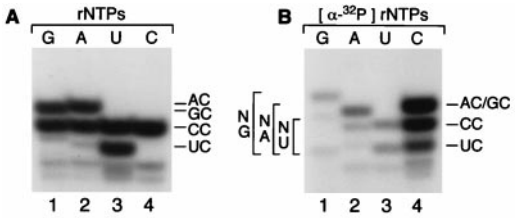

FIG. 2. Template-independent diribonucleotide synthesis by T7 primase. (*A*) Variation of the 5′ nucleotide. Oligonucleotide synthesis reactions contained Mn^{2+} and were performed in the absence of template DNA, with the NTPs indicated and [α-^{32}P]CTP. After synthesis, triphosphate groups were removed by treatment with calf intestinal alkaline phosphatase. Products were separated on a 25% denaturing PAGE gel and autoradiographed. Diribonucleotide species, AC, CC, GC, and UC are identified. (*B*) Variation of the 3′ nucleotide. Oligonucleotide synthesis reactions were performed and analyzed as above, except that reactions contained the indicated [α-^{32}P]NTP and the three other NTPs. Diribonucleotide species NA, NC, NG, and NU are identified.

CC (all lanes), GC (lane 3), and UC (lane 4) equally well. We conclude that the 5′ nucleotide binding site, to which ATP binds in the template-directed reaction (Fig. 1), is nonspecific.

Recognition of Guanine in 5′-GTC-3′. Because any nucleotide may be incorporated in the 5′ position of the dinucleotide, we repeated the reaction in the presence of individual [α-^{32}P]NTPs, the other three NTPs being unlabeled. After phosphatase treatment, the relative incorporation of each nucleotide at the 3′ position can be determined. Relative to NC (100%), the synthesis of NA was 15.2%, NG: 11.6%, and NU: 5.8% (Fig. 2*B*). In contrast with the lack of selectivity for the 5′ nucleotide, primase preferentially incorporates cytosine at the 3′ end of the dinucleotide.

We conclude that the 3′ nucleotide binding site (Fig. 1) preferentially binds CTP. CTP specificity should enable the site to bind CTP before interaction with the template. Instead of the ssDNA template "selecting" the incoming nucleotide, we propose that the bound nucleotide "selects" the template. In short, we propose that CTP, already bound by T7 primase, base pairs with guanine in the primase recognition site 5′-GTC-3′ and so contributes (although not solely) to ssDNA sequence recognition.

Recognition of Thymine in 5′-GTC-3′. In the absence of ssDNA, the 5′ nucleotide binding site is nonspecific and binds any nucleotide. However, in the presence of ssDNA, oligonucleotide synthesis is initiated by primase almost exclusively from ATP (13, 17, 31). This preference indicates that in the presence of ssDNA, either the 5′ nucleotide binding site gains ATP specificity or primase recognizes thymine in the template 5′-GTC-3′, which in turn specifies ATP by base pairing.

To examine these possibilities, oligonucleotide synthesis reactions were performed on templates containing the nucleotide analogs **dU**, **^{5m}C**, and **C** in place of thymine in 5′-G**T**C-3′ (Fig. 3). Primase initiates oligonucleotide synthesis from ATP at both T- and dU-containing sites (lanes 1 and 2) and may also initiate primer synthesis with ATP at a site containing 5-methylcytosine (lane 3) but not cytosine itself (lane 4). Because oligonucleotide synthesis may also be initiated with GTP albeit at a low rate (26), oligonucleotide synthesis reactions were examined with GTP instead of ATP on the templates containing cytosine. Whereas some primer synthesis can be initiated with GTP at the 5′-G**C**C-3′ site, synthesis is far more efficient at the 5′-G^{5m}CC-3′ site (lanes 5 and 6).

The reactions illustrated in Fig. 3 allow the roles in specificity of base pairing, ATP selectivity (in the presence of ssDNA), and thymine recognition by primase to be evaluated individually. The slight but measurable incorporation of GTP opposite cytosine in the template 5′-GCC-3′ (lane 6) demonstrates that base pairing in isolation is sufficient to support

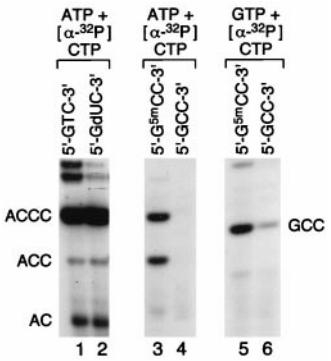

FIG. 3. Oligonucleotide synthesis on templates containing base substitutions for thymine in the primase recognition site 5′-GTC-3′. Oligonucleotide synthesis assays were performed on synthetic oligonucleotide templates with ATP and [α-^{32}P]CTP (lanes 1–4) or GTP and [α-^{32}P]CTP (lanes 5 and 6). The thymine in the recognition site 5′-GTC-3′ was substituted with deoxyuridine (dU), 5-methylcytosine (^{5m}C), or cytosine (C), as indicated (32). After synthesis, products were separated on a 25% denaturing PAGE gel and autoradiographed. Oligonucleotide products are identified.

oligonucleotide synthesis, albeit weakly. In contrast, neither ATP selectivity nor thymine recognition is sufficient individually to support primer synthesis. Thus ATP is not incorporated opposite cytosine in the template 5′-GCC-3′ (lane 4) and no incorporation of GTP is seen opposite thymine in the native template 5′-GTC-3′ (data not shown).

Lanes 3 and 4 in Fig. 3 show the results of two reactions designed to examine the relative contributions of ATP selectivity and thymine recognition toward sequence specificity. There can be no significant base pairing between ATP and 5-methylcytosine, yet primase can incorporate ATP at ^{5m}C-containing sites (lane 3). However, ATP is not incorporated at sites that contain unmodified cytosine (lane 4). Thus the methyl group on C5 of 5-methyl cytosine must be recognized by primase. Extrapolation of this result suggests that primase recognizes the methyl group on C5 of thymine. Similarly, preferential incorporation of GTP at 5-methylcytosine-containing sites over sites containing unmodified cytosine (lanes 5 and 6) implicates the methyl moiety in sequence-specific recognition by primase, from which it can be deduced that the methyl group on C5 of thymine is recognized.

However, methyl group recognition makes only a modest contribution to sequence specificity, which is unmasked solely under suboptimal conditions. Thus, there is little difference between oligonucleotide synthesis initiated with ATP on thymine and uracil-containing templates (Fig. 3, lanes 1 and 2). Therefore, in addition to the major and minor contributions made by base pairing and methyl group recognition, respectively, a further mechanism, as yet undefined, is required to generate the ATP/5′-G**T**C-3′ specificity of primase that is normally observed. This specificity might involve additional contacts between primase and thymine in the template and/or ATP selectivity gained by primase in the presence of ssDNA.

Mutagenic Studies of the Cys$_4$ Zinc Finger of T7 Primase. Because the Cys$_4$ motif is critical for primase activity, we examined its role in direct protein–DNA interactions. Previous studies involving a T3/T7 chimeric primase suggested that the Cys$_4$ loop is involved in sequence-specific recognition of ssDNA (25). Consequently, we altered individual amino acids within the Cys$_4$ loop by *in vitro* mutagenesis and examined their effects on T7 growth and catalytic activities. With the exception of glycine residues, we replaced each of the Cys$_4$ loop residues conserved between T3 and T7 primases with alanine. In some instances, we also made conservative substitutions. The gene 4 mutants were screened for ability to support the growth of T7Δ4–1, a bacteriophage lacking gene 4. T7Δ4–1 is unable to grow in *E. coli* unless complemented with gene 4 protein expressed from a plasmid (29, 33). The locations and identities of the substitutions and their effects on T7Δ4–1 growth are presented in Fig. 4.

All of the altered primases that contained alanine substitutions were able to support the growth of T7Δ4–1, as evidenced by plating efficiencies greater than 0.1 (Fig. 4). In fact, the only substitution to have a major effect on the growth of T7Δ4–1 was the replacement of serine 27 with aspartate. However, because replacement of this same residue with alanine has no effect on T7 viability, we do not believe that serine 27 plays a crucial role in sequence recognition.

Neither altered sequence recognition by primase nor a lowered overall primase activity would be expected to eliminate primer synthesis altogether. Therefore, provided the altered primases are marginally functional, plating efficiency should not be affected. Consequently, we purified several of the altered T7 primases and characterized them biochemically. T7 primase catalyzes the ssDNA-dependent hydrolysis of dTTP, which fuels unidirectional translocation (14). Hence, dTTPase activity provides a quantitative assay of helicase activity. To ensure that none of our alterations affected primase simply by disrupting helicase activity, we analyzed the dTTPase activity of each of the altered primases. All had relatively normal levels of dTTPase activity. (The results are presented in the supplemental material on the PNAS web site, Table 1, column 2).

Each of the purified mutated primases was next examined for its ability to catalyze the synthesis of oligonucleotides. Because sequence specificity might have been changed by individual amino acid substitutions, oligonucleotide synthesis reactions were performed on M13 ssDNA with both ATP + CTP and GTP + CTP. Little or no synthesis was observed from primases S_{27}D, D_{31}A and D_{31}L. All other primases catalyzed oligoribonucleotide synthesis from ATP + CTP. The products of these reactions are illustrated in Fig. 7, which is published as supplemental material on the PNAS web site.

Recognition of the 3′ Cryptic Base. The cryptic base, cytosine, of 5′-GTC-3′ is not involved in base pairing with the primer but is required for initiation of oligonucleotide synthesis (2). This base must therefore be recognized directly by primase. Accordingly, we examined the effect of amino acid substitutions on the recognition of the cryptic cytosine. As a control, oligonucleotide synthesis reactions were carried out with wild-type primase by using synthetic templates that contained the sequences 5′-GT**A**-3′, 5′-GT**C**-3′, 5′-GT**G**-3′,

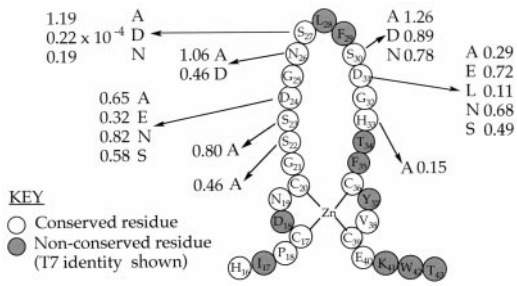

FIG. 4. The Cys$_4$ motif of T7 primase, illustrating position, identity and effect on plating efficiency of single amino acid substitutions in T7 primase. Homology of residues between T3 and T7 primases (11, 34) and identities of individual substitutions are indicated. Numbers represent the efficiency of plating (EOP) of bacteriophage T7Δ4–1, a bacteriophage in which gene 4 is deleted, on *E. coli* cells containing vectors expressing the indicated gene 4. Efficiencies of plating were calculated relative to those obtained for T7Δ4–1 on an *E. coli* strain in which wild-type gene 4 was expressed (EOP = 1.0). Mutations encoding alanine substitutions were generated in the vector pGP4-6$_{G64S10}$. Mutations encoding other substitutions were generated in the vector pET4-6$_{G64S10}$.

and 5′-GT**T**-3′ (Fig. 5*A*). In agreement with previous studies, these data demonstrated that the recognition site 5′-GT**T**-3′ is used minimally (3.1%). The significant use of 5′-GT**A**-3′ and 5′-GT**G**-3′ (12.8% and 15.7%, respectively) was surprising, although RNA-primed DNA synthesis assays used in early studies (13) likely were insufficiently sensitive to detect use of these primase recognition sites.

The oligonucleotide synthesis reactions were repeated with the mutationally altered primases. Alterations to positions Asp_{24} and Asn_{26} yielded primases that deviated from wild-type primase activity only in that recognition of 5′-GT**A**/**G**-3′ was diminished (data not shown). As might be predicted from earlier data, no synthesis was detected from primase $S_{27}D$. However, alterations to two positions, His_{33} and Asp_{31}, significantly affected the sequence specificity of primase.

As shown in Fig. 5*B*, substitution of alanine at position 33 had a striking effect on sequence recognition. Primase $H_{33}A$ is unable to synthesize oligonucleotides on a template that contains the normal recognition site 5′-GT**C**-3′. Surprisingly, however, this altered primase does have activity on the templates 5′-GT**A**/**G**-3′. Thus primase $H_{33}A$ can recognize templates in which the cryptic nucleotide is a purine but is incapable of oligonucleotide synthesis on templates in which the cryptic base is cytosine. We conclude that His_{33} plays an essential role in the recognition of the cryptic cytosine. Underlying cryptic purine recognition is unmasked only when the normal recognition of cytosine is prevented.

The Role of Asp_{31} in Primer Synthesis. The second interesting residue identified in the Cys_4 motif is Asp_{31}. Substitution of Asp_{31} with alanine yields a primase that is inactive on all synthetic templates (Fig. 5*C*, lanes 1–4). Similarly, substitution of Asp_{31} with leucine yields a primase with undetectable activity on these templates (data not shown). In contrast, substitution of Asp_{31} with another acidic residue, glutamate, or an amide, asparagine (Fig. 5*C*, lanes 5–12), yields primases that synthesize primers in proportions similar to wild-type primase (Fig. 5*A*). The inactivities of primases $D_{31}A$ and $D_{31}L$ on all templates suggest that Asp_{31} plays an essential role in primer synthesis. In conjunction, the activity of primase $D_{31}N$ (Fig. 5*C*, lanes 9–12) negates a requirement for negative charge and suggests the importance of a hydrogen bond acceptor at position 31.

Interestingly, unlike the other alterations at position 31, the substitution of serine for aspartate affects the relative use of the synthetic templates. Comparison of the data in Fig. 5 *A* and *C*, lanes 13–16, shows that use of the template 5′-GTC-3′ by primase $D_{31}S$ is significantly reduced. Because cytosine recognition is disrupted by the replacement of Asp (a hydrogen bond acceptor) with Ser (which can act either as a hydrogen bond donor or acceptor), we examined cytosine itself for moieties that might hydrogen bond with Asp in the wild-type enzyme. Cytosine contains only one such moiety: the amino group on C4. To investigate the role of this amino group, we carried out oligonucleotide synthesis reactions with both wild-type and the $D_{31}S$ primases on a template containing the sequence 5′-GT^{4H}C-3′. In agreement with earlier results (32), removal of the amino group had little effect on wild-type primase activity (Fig. 5*D*). However, removal of this amino group stimulated primer synthesis 2.6-fold by the substituted primase $D_{31}S$ (Fig. 5*D*).

We conclude that the amino group on C4 of cytosine inhibits oligonucleotide synthesis by primase $D_{31}S$. Inhibition most likely results from unfavorable interaction(s) involving the amino group of cytosine. Most probably, substitution of Ser for Asp_{31} leads to a conformational change within primase that in turn causes a steric conflict between primase and the amino group on C4 of cytosine. There is, however, an alternative explanation. In the wild-type primase, Asp_{31} might also form a hydrogen bond with the amino group of cytosine. This electrostatic attraction could be converted to repulsion by the substitution of Ser for Asp_{31}, but might be maintained by the substitutions $D_{31}E$ and $D_{31}N$.

DISCUSSION

We have examined the role of the Cys_4 zinc motif, or "zinc ribbon," in sequence-specific recognition of ssDNA. We now summarize our findings and consider their relevance to other single Cys_4 motif-containing proteins.

Sequence-Specific Recognition of ssDNA by T7 Primase. At least three independent regions of T7 primase contribute to the sequence-specific recognition of 5′-GTC-3′. We propose that the primase 3′ nucleotide binding site (Fig. 1) initially binds CTP, which in turn, by base pairing, recognizes G of 5′-GTC-3′. Meanwhile, the Cys_4 motif, using residue His_{33}, contributes to the recognition of C in 5′-GTC-3′. Our results also suggest that primase recognizes the methyl group on C5 of the central nucleotide, T, of 5′-GTC-3′ by an unknown mechanism. Because we have demonstrated that ATP specificity of the 5′ nucleotide binding site is gained only in the presence of ssDNA, we suggest that this specificity results from base pairing between ATP and T in 5′-GTC-3′. However, Asp_{31} may also contribute to ATP recognition, perhaps by hydrogen bonding with ATP in the 5′ nucleotide binding site (see below).

The Mechanism of Oligonucleotide Synthesis by T7 Primase. *E. coli* RNA polymerase synthesizes RNA by a mixture of monotonous and "inchworm-like" movements (35). In monotonous movement, the entire RNA polymerase moves forward by one nucleotide at a time. During "inchworm-like" movement, the front end of the enzyme remains anchored while the catalytic site continues to polymerize nucleotides (35). *E. coli* RNA polymerase is thought to use the "inchworm" model of elongation during initiation of RNA synthesis, and it

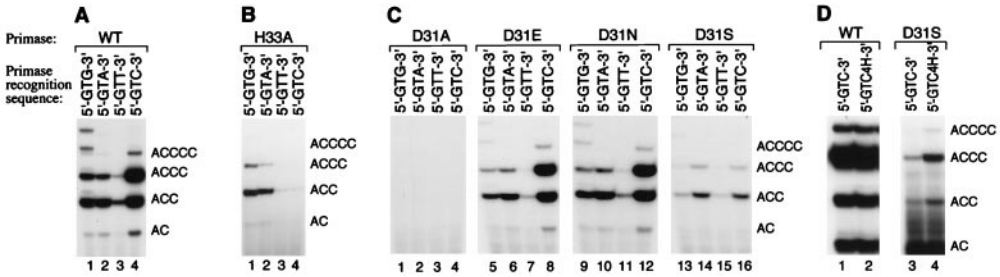

FIG. 5. Oligonucleotide synthesis by wild-type primase and primases genetically altered at positions 31 and 33 on templates containing base substitutions for the cryptic cytosine. Oligonucleotide synthesis assays were performed with ATP and [α-^{32}P]CTP on synthetic ssDNA oligonucleotide templates containing primase recognition sites of sequence 5′-GCN-3′, where N is A, C, G, or T, as indicated (Fig. 5 *A–C*) or 5′-GCC^{x}-3′, where C^{x} is either C or ^{4H}C, as indicated (Fig. 5*D*). After synthesis, products were separated on a 25% denaturing PAGE gel and autoradiographed. Oligonucleotide products are identified. (*A*) Wild-type primase. (*B*) Primase $H_{33}A$. (*C*) Primases $D_{31}A$, $D_{31}E$, $D_{31}N$ and $D_{31}S$. (*D*) Wild-type and $D_{31}S$ primases (for comparison with $D_{31}S$, the products of wild-type primase were diluted 10-fold before electrophoresis).

is possible that *E. coli* primase uses a similar mechanism. In each case, if the initiating nucleotide, ATP, is crosslinked to the enzyme, nucleotide extension can continue for four or five bases in the case of primase (36) or for eight or nine bases in the case of RNA polymerase (37). We believe that an analogous mechanism may operate in T7 primase, in which ATP is held in the 5′ nucleotide binding site, while elongation is performed by the 3′ nucleotide binding site.

In accordance with this model, our finding that the 3′ nucleotide binding site is CTP specific may account for hitherto unexplained results. During "pseudotemplating" an additional cytosine, not specified by the template, is added to the 3′ end of a primer (2). Furthermore, on long polyguanidine templates, long poly-C-containing oligonucleotides may be synthesized (38). Although T7 primase can synthesize long oligonucleotides (25, 38), the incorporation of cytidylate is favored greatly, and only those with high cytosine contents are extended by T7 DNA polymerase (13, 16, 17, 38). We propose that this bias also results from the high affinity of the 3′ nucleotide binding site: a primer containing cytosine at its 3′ end is held more tightly by T7 primase and is therefore passed more efficiently to T7 DNA polymerase.

The Cys_4 Motif of T7 Primase: Interactions with ssDNA and ATP. Our results demonstrate that two residues within the Cys_4 motif, Asp_{31} and His_{33}, play major roles in sequence recognition. We have suggested the importance of a hydrogen bond acceptor at position 31. As such, Asp_{31} might simply be involved in the structural integrity of primase. Alternatively, Asp_{31} might accept a hydrogen bond from thymine in the site 5′-GTC-3′ or from the incoming 5′ nucleotide, ATP. Of these, we consider a hydrogen bond between thymine and Asp_{31} unlikely because thymine has only one hydrogen bond donor, which is required for base pairing with ATP. In contrast, after base pairing ATP still has a hydrogen bond donor available, the second proton of the C6 amino group. Studies with *E. coli* primase have shown that the initiating nucleotide, ATP, can be crosslinked both to residues within the nucleotide binding site and to residues in its $CysHisCys_2$ zinc finger (36). We therefore propose a hydrogen bond between the amino group on C6 of ATP and the carboxyl group of Asp_{31}.

The role of the second important residue, His_{33}, is more easily defined. We have demonstrated that His_{33} is essential for cryptic cytosine recognition. Substitution of alanine for His_{33} might alter the specificity of primase simply by making space for a larger purine base at this position in the template. More likely, the substitution may directly eliminate an interaction with cytosine. Inasmuch as earlier studies had shown that N3 of the cryptic cytosine is essential for recognition (32), we propose that this nitrogen hydrogen bonds with an imidazole proton of His_{33}.

Because the Cys_4 motifs of T7 primase and TFIIS share a common consensus sequence (9), a molecular model of the T7 primase Cys_4 domain based on the Cys_4 domain of TFIIS (PDB:1TFI) was constructed (Fig. 6*A*). Unfortunately, attempts to generate a similarly minimized segment of an RNA–DNA helix comprising 5′-GTC-3′ (DNA) and 5′-AC-3′ (RNA) were unsuccessful: energy minimization merely corrupted the base pairing between the two strands of nucleic acid. We therefore constructed a physical replica of residues Asp_{24}-His_{33} of the energy-minimized Cys_4 motif and a DNA/RNA helix of sequence 5′-GTC-3′/5′-AC-3′. We then docked the two models together by connecting Asp_{31} to the amino groups of ATP and dCMP and the imidazole proton of His_{33} to N3 of cytosine with hydrogen bonds. Rotation around the bonds between the α, β, and γ carbons of each side chain permitted parallel orientation of all bases in the RNA/DNA strands. The resulting model of the Cys_4 motif/DNA/RNA complex is illustrated in cartoon format (Fig. 6*B*). Although highly speculative, the model is consistent with observed results. For example, in addition to the proposed hydrogen bonds, Ser_{27} is in sufficiently close proximity to form a hydrogen bond with the sugar-phosphate backbone between bases G and T of the trinucleotide. Consistent with such an interaction, we find that an acidic residue cannot be tolerated at position 27. If Ser_{27} does form a hydrogen bond with the sugar-phosphate backbone, then substitution $S_{27}D$, which destroyed primase activity, would lead to a repulsion between the protein and ssDNA. Two hydrophobic interactions are also suggested by the model: a van der Waals interaction between the methyl group of thymine and Leu_{28} and an intercalation of

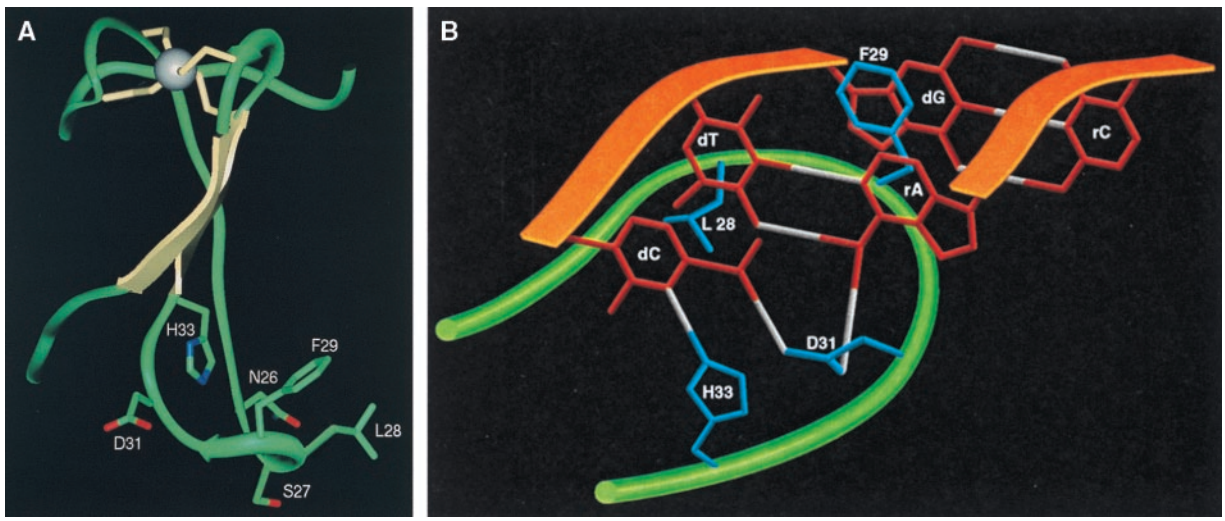

FIG. 6. Models of the Cys_4 zinc-binding motif of T7 primase. (*A*) Model based on the zinc ribbon of TFIIS. Coordinates of the NMR ensemble of TFIIS (nucleic acid binding domain, *Homo sapiens*, Protein Data Bank ID code 1TFI) (9) were used to model the Cys_4 motif of bacteriophage T7 primase. Initially, the first nine conformers of TFIIS were aligned and structurally conserved regions (SCRs) were identified by automatic structure alignment [INSIGHTII ver. 95.0 Homology module, default parameters (43)]. The amino acid sequence of TFIIS was sequentially aligned, and coordinates of the T7 primase were built from that of TFIIS by using coordinates previously determined by SCRs. The gaps were consequentially closed by using coordinates of one of the conformers of TFIIS. Energy minimization was then applied by using the Amber forcefield within the INSIGHTII ver. 95.0 Discover module (44). Fifty steps of steepest decent energy minimization were used, followed by 200 steps of conjugate gradient energy minimization. (*B*) Illustration of proposed interaction between residues Asp_{24}–His_{33} of the T7 primase Cys_4 motif, the ssDNA recognition sequence, 5′-GTC-3′, and the diribonucleotide AC. Peptide and phosphodiester backbones are represented as green and orange ribbons, respectively.

Phe_{29} between bases T and G. These hydrophobic residues are functionally conserved in the Cys_4 motif of bacteriophage T3 primase (34). The proposed contacts are all formed with the major groove of the RNA/DNA helix.

Is There a Conserved Role for the Single Cys_4 Motif? Many examples of Cys_2His_2 zinc fingers are known and the role of these "mini-domains," first proposed by Miller *et al.* (39), has been analyzed from structures of protein–DNA complexes (e.g., 40). In essence, one zinc finger contacts three bases of double-stranded DNA. Protein–DNA contacts are made by residues in known positions within the motif and consecutive zinc fingers may bind successive trinucleotides of double-stranded DNA (40). In comparison, single Cys_4 "zinc ribbons" are structurally dissimilar to any other type of DNA-binding domain (3). Although the Cys_4 motif of T7 primase also mediates the recognition of three bases of DNA, it is ssDNA rather than double-stranded DNA. Moreover, existing characterizations of single Cys_4 motifs suggest that the motif contributes to protein–DNA interactions, rather than making those interactions *per se*. For example, in TFIIS both the Cys_4 motif and a contiguous region of the protein are required for protein–ssDNA interaction (4). Similarly, although the two subunits of TFIIE bind ssDNA in concert, the large subunit, which contains a Cys_4 motif, facilitates binding to ssDNA by the small subunit, rather than interacting with the DNA itself (41). Again, adenovirus E1A protein is thought to activate transcription by interacting with sequence-specific transcription factors, an activity for which both its Cys_4 motif and a contiguous region are essential (42).

Given that single TFIIS-like Cys_4 motifs are found in functionally related proteins that bind ssDNA, we suggest that Cys_4 "zinc ribbons" have a conserved role. Our results suggest that the Cys_4 motif of T7 primase recognizes only one base of the trinucleotide 5′-GTC-3′ directly, and that it contributes to the recognition of a second base. However, the third base appears to be recognized by a different domain of T7 primase: the 3′ nucleotide binding site. We postulate that rather than recognizing a trinucleotide of ssDNA directly, the Cys_4 motif instead recognizes one or two bases of ssDNA and so stabilizes DNA recognition by other domains of T7 primase. By extrapolation, we propose that the single Cys_4 motif may act similarly in other proteins.

We are grateful to Benjamin B. Beauchamp, Joonsoo Lee, and Stanley Tabor for providing purified wild-type T7 gene 4 primase and bacteriophage T7Δ4–1. We thank Dr. Ian O. Sutherland (University of Liverpool, Liverpool, U.K.) and Dr. Thomas Ellenberger (Harvard Medical School) for helpful discussions and David Frick, Leo Guo, Ingrid Richardson, and Michael Sawaya for critical readings of the manuscript. This investigation was supported by grant number AI-06045 from the United States Public Health Service and grant number NP-1–28 from the American Cancer Society to C.C.R. and by a grant from the Pharmaceutical Institute, Aston University, Birmingham, U.K. to A.V.H..

1. Cha, T. A. & Alberts, B. M. (1986) *J. Biol. Chem.* **261,** 7001–7010.
2. Mendelman, L. V. & Richardson, C. C. (1991) *J. Biol. Chem.* **266,** 23240–23250.
3. Qian, X., Jeon, C. J., Yoon, H. S., Agarwal, K. & Weiss, M. A. (1993a) *Nature (London)* **365,** 277–279.
4. Qian, X., Jeon, C. J., Yoon, H. S., Agarwal, K. & Weiss, M. A. (1995) *Nature (London)* **376,** 279.
5. Agarwal, K., Baek, K. H., Jeon, C. J., Miyamoto, K., Ueno, A. & Yoon, H. S. (1991) *Biochemistry* **30,** 7842–7851.
6. Bernstein, J. A. & Richardson, C. C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 396–400.
7. Ilyina, T. V., Gorbalenya, A. E. & Koonin, E. V. (1992) *J. Mol. Evol.* **34,** 351–357.
8. Feaver, W. J., Henry, N. L., Bushnell, D. A., Sayre, M. H., Brickner, J. H., Gileadi, O. & Kornberg, R. D. (1994) *J. Biol. Chem.* **269,** 27549–27553.
9. Qian, X., Gozani, S. N., Yoon, H. S., Jeon, C. J., Agarwal, K. & Weiss, M. A. (1993b) *Biochemistry* **32,** 9944–9959.
10. Zhu, W., Zeng, Q., Colangelo, C. M., Lewis, L. M., Summers, M. F. & Scott, R. A. (1996) *Nat. Struct. Biol.* **3,** 122–124.
11. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535.
12. Hingorani, M. M. & Patel, S. S. (1993) *Biochemistry* **32,** 12478–12487.
13. Tabor, S. & Richardson, C. C. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 205–209.
14. Matson, SW. & Richardson, C. C. (1983) *J. Biol. Chem.* **258,** 14009–14016.
15. Scherzinger, E., Lanka, E. & Hillenbrand, G. (1977) *Nucleic Acids Res.* **4,** 4151–4163.
16. Romano, L. J. & Richardson, C. C. (1979a) *J. Biol. Chem.* **254,** 10483–10489.
17. Fujiyama, A., Kohara, Y. & Okazaki, T. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 903–907.
18. Bird, L. E., Hakansson, K., Pan, H. & Wigley, D. B. (1997) *Nucleic Acids Res.* **25,** 2620–2626.
19. Notarnicola, S. M. & Richardson, C. C. (1993) *J. Biol. Chem.* **268,** 27198–27207.
20. Patel, S. S., Hngorani, M. M. & Ng, W. M. (1994) *Biochemistry* **33,** 7857–7868.
21. Notarnicola, S. M., Park, K., Griffith, J. D. & Richardson, C. C. (1995) *J. Biol. Chem.* **270,** 20215–20224.
22. Notarnicola, S. M., Mulcahy, H. L., Lee, J. & Richardson, C. C. (1997) *J. Biol. Chem.* **272,** 18425–18433.
23. Frick, D. N., Baradaran, K. & Richardson, C. C. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7957–7962.
24. Mendelman, L. V., Beauchamp, B. B. & Richardson, C. C. (1994) *EMBO J.* **13,** 3909–3916.
25. Hine, A. V. & Richardson, C. C. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12327–12331.
26. Kusakabe, T. & Richardson, C. C. (1996) *J. Biol. Chem.* **271,** 19563–19570.
27. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual,* 2nd Ed. (Cold Spring Harbor Lab. Press, Plainview, New York), pp. A9–A12.
28. Mendelman, L. V., Notarnicola, S. M. & Richardson, C. C. (1993) *J. Biol. Chem.* **268,** 27208–27213.
29. Tabor, S. (1987) Ph.D. thesis (Harvard University Medical School, Boston, MA).
30. Kusakabe, T. & Richardson, C. C. (1997a) *J. Biol. Chem.* **272,** 12446–12453.
31. Romano, L. J. & Richardson, C. C. (1979b) *J. Biol. Chem.* **254,** 10476–10482.
32. Mendelman, L. V., Kuimelis, R. G., McLaughlin, L. W. & Richardson, C. C. (1995) *Biochemistry* **34,** 10187–10193.
33. Mendelman, L. V., Notarnicola, S. M. & Richardson, C. C. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10638–10642.
34. Beck, P. J., Gonzalez, S., Ward, C. L. & Molineux, I. J. (1989) *J. Mol. Biol.* **210**, 687–701.
35. Nudler, E., Goldfarb, A. & Kashlev, M. (1994) *Science* **265,** 793–796.
36. Mustaev, A. A. & Godson, G. N. (1995) *J. Biol. Chem.* **270,** 15711–15718.
37. Mustaev, A., Kashlev, M., Zaychikov, E., Grachev, M. & Goldfarb, A. (1993) *J. Biol. Chem.* **268,** 19185–19187.
38. Kusakabe, T. & Richardson, C. C. (1997b) *J. Biol. Chem.* **272,** 5943–5951.
39. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
40. Kim. C. A. & Berg, J. M. (1996) *Nat. Struct. Biol.* **3,** 940–945.
41. Kuldell, N. H. & Buratowski, S. (1997) *Mol. Cell. Biol.* **17,** 5288–5298.
42. Geisberg, J. V., Lee, W. S., Berk, A. J. & Ricciardi, R. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2488–2492.
43. Homology User Guide (1995) (Biosym/Molecular Simulations Inc., San Diego, CA).
44. Discover User Guide (1995) (Biosym/Molecular Simulations, Inc., San Diego, CA).