Crystal Structure of Bacteriophage T4 5′ Nuclease in Complex with a Branched DNA Reveals How Flap Endonuclease-1 Family Nucleases Bind Their Substrates

Bacteriophage T4 RNase H, a flap endonuclease-1 family nuclease, removes RNA primers from lagging strand fragments. It has both 5′ nuclease and flap endonuclease activities. Our previous structure of native T4 RNase H (PDB code 1TFR) revealed an active site composed of highly conserved Asp residues and two bound hydrated magnesium ions. Here, we report the crystal structure of T4 RNase H in complex with a fork DNA substrate bound in its active site. This is the first structure of a flap endonuclease-1 family protein with its complete branched substrate. The fork duplex interacts with an extended loop of the helix-hairpin-helix motif class 2. The 5′ arm crosses over the active site, extending below the bridge (helical arch) region. Cleavage assays of this DNA substrate identify a primary cut site 7-bases in from the 5′ arm. The scissile phosphate, the first bond in the duplex DNA, is 5′ adjacent to the 5′ arm, and lies above a magnesium binding site. The less ordered 3′ arm reaches toward the C and N termini of the enzyme, which are binding sites for T4 32 protein and T4 45 clamp, respectively. In the crystal structure, the scissile bond is located within the double-stranded DNA, between the first two duplex nucleotides next to the 5′ arm, and lies above a magnesium binding site. This complex provides important insight into substrate recognition and specificity of the flap endonuclease-1 enzymes.

The flap endonuclease-1 (FEN-1) family nuclease family is conserved in sequence and structure from bacteriophage to humans. These nucleases play essential roles in DNA replication by removing the RNA primers from lagging strand fragments. In addition, FEN-1-related nucleases are important in long-patch base excision repair and in maintenance of genomic stability. Homozygous knockouts of FEN-1 in mice are lethal to embryos (for review, see Refs. 1–3).

Like other family members, bacteriophage T4-encoded RNase H shows 5′ to 3′ exonuclease activity on either RNA/DNA or DNA/DNA duplexes and endonuclease activity on either flap or fork DNA structures (4, 5). T4 rnh deletion mutants give no phage production and accumulate unligated, lagging strand fragments in Escherichia coli hosts with defective polymerase I 5′ nuclease (6). In addition, the mutants are hypersensitive to UV irradiation and anti-tumor agents (7).

FEN-1 family nuclease activities are modulated by interactions with DNA replication clamps (e.g. eukaryotic proliferating cell nuclear antigen and T4 gene 45 clamp) and single-stranded DNA-binding proteins (e.g. eukaryotic replication protein A and T4 gene 32 protein) (5, 8–12). Human FEN-1 nuclease is also stimulated by interactions with the Werner (13, 14) and Bloom (15) syndrome helicases as well as the Rad9-Rad1-Hus1 (9–1–1) checkpoint clamp (16).

T4 RNase H 5′ nuclease removes a short oligonucleotide (1–4 bases) each time it binds its substrate. T4 32 protein, binding on single-stranded DNA behind the nuclease, increases its processivity so that about 10–50 short oligonucleotides are removed in a single binding event (5). On nicked substrates, T4 RNase H is stimulated by the T4 replication clamp but not by 32 protein (9). The interaction of T4 RNase H with 32 protein is largely responsible for the maturation of lagging strand fragments in the T4 replication system in vitro and T4 phage production in vivo (9). The primers are normally removed, whereas the single-stranded DNA covered by 32 protein is still between the nuclease and the polymerase that extends the adjacent fragment. The clamp likely serves a backup function, allowing the nuclease to be loaded on nicked molecules that are formed when polymerase completes synthesis of the next fragment but before RNase H removes the primer. The T4 mechanism is different from current models for the maturation of lagging strand fragments in eukaryotes. In the latter, strand displacement synthesis by polymerase δ creates flaps that are removed mainly by the FEN-1 flap endonuclease, a reaction stimulated by the proliferating cell nuclear antigen clamp (17–20).

Our structure of native T4 RNase H (PDB code 1TFR) revealed an active site composed of six highly conserved Asp...
residues plus two hydrated Mg$^{2+}$ ions (21). This structure shows strong similarity to those in other members of the FEN-1 family (22–28). In the present report, we describe the 3.0-Å x-ray analysis of an active site mutant of T4 RNase H co-crystallized with a fork DNA ligand. This is the first structure of a FEN-1 enzyme containing a branched DNA substrate in its active site. No major conformational changes were detected in the protein structure upon binding of the DNA molecule. The crystal structure of the enzyme-substrate complex clearly delineates the arrangement of T4 RNase H, T4 32 protein, and the T4 45 clamp on the lagging strand and provides insight into the range of specificities displayed by individual members of the FEN-1 family of proteins.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Recombinant native T4 RNase H (pNN2202) and the inactive D132N mutant (pNN2202-D132N) (29) were expressed in *E. coli* BL21(DE3) pLysS. Cells were grown in 6 liters of Luria-Bertani broth with 45 μg/ml ampicillin to an A$_{600}$ of 0.6 (37 °C), induced with 1 mm isopropyl β-D-thiogalactopyranoside for 3 h, harvested by centrifugation, and pellets were stored at −20 °C. Cells were suspended in 10 ml of lysis buffer (50 mM Tris-HCl, pH 7.5, 0.2 mM NH$_4$Cl, 10 mM MgCl$_2$, 5% glycerol, 2 mM dithiothreitol, 1 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride, and 0.3% polyethyleneimine) per gram of cell pellet, lysed using sonication, and centrifuged (20,000 × g, 4 °C, 30 min). Conductivity of the samples was reduced by dilution using buffer (50 mM Tris-HCl, pH 7.5) for each chromatographic step. Lysates were loaded onto a 50-ml SP-Sepharose column (SP Fast Flow, Amersham Biosciences) equilibrated with SP buffer A (50 mM Tris-HCl, pH 7.5, 100 mM NH$_4$Cl, and 1% glycerol) and then eluted with SP buffer B (SP buffer A containing 0.75 M NaCl). Pooled fractions were loaded onto a 10-ml ceramic hydroxyapatite column (Bio-Rad Macro-Prep Type II) equilibrated with HA buffer A (25 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 1% glycerol) and then eluted with HA buffer B (HA buffer A plus 1 M (NH$_4$)$_2$SO$_4$). Pooled HA fractions were loaded onto a 25-ml POROS HS column (Applied Biosystems) equilibrated with SP buffer A and eluted using SP buffer B; the protein was displaced at ~350 mM NaCl. The purified proteins were concentrated by centrifugation (Millipore Amicon Ultra, 10,000 molecular weight cutoff), cryo-protected with 20% glycerol, and stored as aliquots of 1 ml at −80 °C.

**Preparation of Oligonucleotides**—Oligonucleotides containing 18 and 24 bases to act as fork DNA substrate were designed according to previously reported sequences (30). High performance liquid chromatography-purified oligonucleotides (IDT, Coralville, IA) were dissolved in buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 100 mM NaCl) to a final concentration of 5 mM. Annealed fork substrates were prepared by first mixing 1:1 molar ratios of two partially complementary DNA strands in 0.5-ml Eppendorf tubes. These tubes were placed in hot water (95 °C), slow-cooled to room temperature overnight, moved to 4 °C, and then stored at −20 °C. Samples of fork DNA substrates used in the crystallization experiments were a 12-base complementary duplex, a 6-base 5’ arm, and a 12-base 3’ arm (Scheme 1).

**Crystallization**—Purified D132N T4 RNase H was dialyzed at 4 °C against 25 mM Bis-Tris, pH 6.5, 2 mM EDTA, 150 mM NH$_4$Cl, concentrated to 20 mg/ml (Millipore Amicon Ultra, 10,000 molecular weight cutoff), and mixed in an equimolar ratio with fork DNA substrate (10 mg/ml protein). Sparse matrix screening at 23 °C (Emerald Biostructures, Inc.) yielded crystals that were optimized by hanging drop vapor diffusion of droplets composed initially of 2 μl of protein, 4 μl of water, and 2 μl of the well solution. The latter consisted of 0.1 M Tris-HCl, pH 7.5, 18% polyethylene glycol 2000 monomethyl ether (MME), and 20% glucose.

**Data Collection and Processing**—Crystals for data collection were flash-cooled in liquid nitrogen, and diffraction data were collected at 100 K on a Rigaku FR-E High Brilliance x-ray generator (copper anode) interfaced with a Saturn 92 CCD detector (The Ohio Macromolecular Crystallography Consortium, The University of Toledo, Toledo, OH). Analysis of the x-ray data indicated that the enzyme-substrate complex crystallized in the orthorhombic space group P2$_1$2$_1$2$_1$, with cell dimensions $a = 68.16$ Å, $b = 84.99$ Å, and $c = 89.29$ Å (Table 1). Data, which extended to 3.0 Å resolution, were processed with CrystalClear (Version 1.3.5, Rigaku MSC). Crystal statistics are presented in Table 1. Attempts to collect higher resolution data were unsuccessful. The Matthews coefficient of 2.7 and the solvent content of 53.4% indicated that the crystallographic asymmetric unit
consisted of one D132N RNase H molecule and one bound 18/24 fork DNA ligand (32).

Structure Determination and Refinement—The structure was solved by molecular replacement with the program AMoRe (33) using native T4 RNase H (PDB code 1TFR) as the starting model (21). Nucleic Acid Builder was used to generate the DNA coordinate file with the appropriate sequence. A 6-bp DNA duplex was first manually positioned in the initial electron density map using the program O (34). Several rounds of model building and refinements allowed us to fit the nucleotides of the fork DNA substrate, typically adding a few nucleotides after each refinement. Refinements were carried out with CNS (35) followed by Refmac5 (36). The \( \psi/\phi \) angles in the Ramachandran plot of the final protein model show the following distribution (using PROCHECK (37)): 81.5% most favored, 16.3% allowed, 1.8% generously allowed, and 0.4% disallowed (1 residue, Lys-128, see “Results”). The complete 12-bp duplex, five bases of the 5’ arm (PDB chain C) and five bases of the 3’ arm (PDB chain D) were visible in the final model. Disorder prevented modeling of the remaining bases. Molecular figures of the enzyme and its DNA ligand were displayed and rendered with the program PyMOL. The enzyme-substrate complex was superimposed on the native T4 RNase H and on the Afu FEN-1 structures (PDB).

### Table 1

Data collection and refinement statistics

| PDB accession code | 2IHN |
|--------------------|------|
| **Data collection** |      |
| Space group        | P2_1 |
| Cell dimensions    |      |
| a, b, c (Å)        | 68.2, 85.0, 89.3 |
| \( \alpha, \beta, \gamma \) (°) | 90.0, 90.0, 90.0 |
| Wavelength (Å)     | 1.5418 |
| Resolution (Å)     | 38.4-3.0 (3.1-3.0) |
| \( R_{merge} \)     | 9.1 (32.8) |
| I/\( I_{ave} \)     | 12.7 (3.9) |
| Observed reflections| 41,111 |
| Unique reflections  | 10,142 |
| Completeness (%)   | 99.2 (99.8) |
| Redundancy         | 3.8 (3.7) |
| **Refinement**     |      |
| Resolution (Å)     | 38.4-3.0 |
| \( R_{work}/R_{free} \) | 22.2 (34.4)/28.2 (39.2) |
| No. atoms (non-hydrogen) |    |
| Protein            | 2,459 |
| DNA                | 696 |
| Water              | 12 |
| B-Factors          |      |
| Protein            | 34.4 |
| DNA                | 49.6 |
| Water              | 30.7 |
| Root mean square deviations |    |
| Bond lengths (Å)   | 0.021 |
| Bond angles (°)    | 2.37 |

| Ramachandran plot  |      |
| Most favored (%)   | 81.5 |
| Allowed (%)        | 16.3 |
| Generously allowed | 1.8  |
| Disallowed (%)     | 0.4  |

Numbers with the associated arrows identify the sites of DNA cleavage. DNA size markers (M) are snake venom phosphodiesterase (PDE) or DNA polymerase I large Klenow fragment (KF) digests of the 5' 32P-labeled 18-mer. C, hydrolysis of single and double flap DNA by 0.1–25 nM wild type T4 RNase H. The substrates had identical sequences except for the extra unpaired 3’ nucleotide in the double flap.

**FIGURE 1.** Binding and digestion of linear, fork, and flap DNA by T4 RNase H. For all panels, the fork DNA substrates are shown on top, and the 5' 32P label is indicated with an asterisk (*). A, mobility shift assays with the D132N mutant of T4 RNase H (see “Experimental Procedures”). The substrates are composed of a 12-mer double-stranded DNA duplex. The labels correspond to the lengths of the corresponding arms. B, hydrolysis of the 18/24 fork by wild type T4 RNase H. T4 RNase H concentration increased from 0.25 to 20 nM. Numbers with the associated arrows identify the sites of DNA cleavage. DNA size markers (M) are snake venom phosphodiesterase (PDE) or DNA polymerase I large Klenow fragment (KF) digests of the 5' 32P-labeled 18-mer. C, hydrolysis of single and double flap DNA by 0.1–25 nM wild type T4 RNase H. The substrates had identical sequences except for the extra unpaired 3’ nucleotide in the double flap.
code 1RXW) (23) using the secondary-structure matching superposition program inside COOT (38). Atomic coordinates and structure factors have been deposited.

RESULTS

T4 RNase H DNA Complexes and Nuclease Activity—Fig. 1A illustrates the analyses of linear and fork DNA substrates by electrophoretic mobility shift assay to identify the smallest DNA fragment that could form a tight complex with T4 RNase H. Successful candidates contained a 12-base duplex with varying lengths of non-complementary 5’ and 3’ arms. Gel shift assays indicated that a 6-base 5’ arm and a 12-base 3’ arm (the 18/24 fork substrate) are required for formation of stable binary complexes. Binding is not improved by the introduction of a 3’ duplex arm (data not shown) or the lengthening of the 3’ arm to 18 bases or an extension of the 5’ arm to 12 bases. Binding is abrogated in the absence of a 5’ arm.

Hydrolysis of the 18/24 fork DNA primarily released a 7-mer single-stranded DNA fragment from the 5’ arm, with the cut site located one base into the duplex region near the branch point (Fig. 1B). Lower amounts of the 5- and 8-mers and traces of the 6-mer were also found. In contrast to the FEN-1 enzymes, which exhibit a preference for a double-flap DNA substrate with a short 5’ arm and a single 3’ overhanging base at the branch point (39), T4 RNase H normally shows preference for single-flap and pseudo-Y (fork) substrates (5). Although catalytic activity seems opposite to what is usually observed for FEN-1 enzymes, the cleaving specificity of T4 RNase H on these three substrates is the same as that displayed by FEN-1 enzymes (30, 40). In our assays (Fig. 1C), T4 RNase H released mostly 19- and 21-base nucleotides from the single flap substrates, cleaving the flap strand on either side of the branch point but not at the branch point (20-base nucleotide product). Activity on double-flap substrate released 21 bases (one base into the duplex region), creating a nicked product identical to those generated by FEN-1 enzymes (Fig. 1C). However, the amount of hydrolysis was much lower for this substrate compared with the two other substrates (fork and single flap).

Structure of T4 RNase H Bound to a Fork DNA Substrate—The two hydrated magnesium ions of the native enzyme, Mg1 and Mg2, are coordinated through water to the cluster of conserved acidic residues, with Mg1 having inner-sphere coordination to Asp-132 (21). Any Asp to Asn mutation of residues coordinated to Mg2 (Glu-130, Asp-157, and Asp-200) only reduces the nuclease activity of T4 RNase H (29).

Early attempts to co-crystallize chelated native enzyme with DNA substrates failed to produce any crystals. However, using the D132N mutation to eliminate catalytic activity, crystals were obtained of the enzyme in complex with 18/24 branched DNA substrate. Data collection and refinement statistics are shown in Table 1. These statistics are comparable with other crystal structures of protein-DNA complexes that have been solved at similar resolution. Initial phases were obtained using molecular replacement of the protein model alone. A stereo image of the initial Fo – Fe electron density map is presented in Fig. 2A showing the density in the region of the DNA using phases obtained from the protein alone. Initially, a double-stranded 6-mer was manually positioned in this electron density with additional nucleotides added as phases improved. D132N T4 RNase H consists of residues 1–305, of which residues 91–96 are missing in the bridge region of the refined structure. The 18/24 fork DNA is modeled as a 17/17 fork DNA, with a double-stranded 12-mer and two 5-mer single-stranded arms. The final Fo – Fe electron density map is presented in Fig. 2B. The first base of the 5’ arm and the last seven bases of the 3’ arm are disordered. The entire 12-mer duplex belongs to the B-form of DNA with all Watson-Crick base pairs intact, including the base pair at the branch point. The 5’ arm traverses the active site cleft and continues in a direction perpendicular to the 12-mer duplex helical axis (Fig. 2C), whereas the 3’ arm follows a path parallel to the duplex region toward the C terminus of the protein (Fig. 2D and supplemental movie 2D associated with Fig. 2D).

In its native form, T4 RNase H can be divided into a large αβ subdomain and a small α-helical subdomain. A short disordered bridge region extends above the cleft that contains the active site and two bound Mg2+ ions (21). The new model of the T4 RNase H in a complex with fork DNA provides critical information for understanding the structure-specific DNA recognition by the FEN-1 family of enzymes (Fig. 2, C and E). In Fig. 2F, the amino acid sequence of T4 RNase H is aligned with the secondary structure to highlight the interactions with the DNA substrate. These interactions are examined in more detail in Fig. 3A.

A helix-hairpin-helix (HHH) motif was previously described for T4 RNase H (24). HHH is a non-sequence-specific DNA

![FIGURE 2. Structure of T4 RNase H bound to branched DNA.](image-url)
binding motif consisting of two antiparallel α helices connected by a four-residue hairpin (type II β turn) (41). The 5’ to 3’ nuclease have extended insertions in the hairpin region (42), modifying the motif name to H3TH (24). In T4 RNase H, the corresponding structure is designated the HhH motif class 2 (HhH2) (European Bioinformatics Institute, InterPro accession number IPR008918). Sites in the fork DNA recognized by the enzyme include the 12-mer duplex region bound to the HhH2 motif (Fig. 3), the branch point, the 5’ arm traversing the active site under the bridge (Fig. 3C), and the 3’ arm extending from the branch point to the C terminus of the enzyme. Interactions of the enzyme with the duplex region and the 3’ arm segment of DNA involve mainly polar residues. In contrast, the branch point and the 5’ arm near the active site tend to make more hydrophobic contacts with the protein.

Interactions of the HhH2 Motif with the Duplex Segment of DNA—T4 RNase H interacts with double-stranded DNA at two specific sites. The HhH2 region, helices H7 through H8, impinges on the duplex DNA section of the 3’ arm (chain D) (Fig. 3, A and B), and His-174 closely approaches the branch point (Fig. 3C). As in the hallmark interactions of an HhH fold (42), the peptide amide groups of residues Thr-221, Met-224, and Thr-226 form hydrogen bonds with the phosphate oxygen atoms of D6Gua, D5Thy, and D4Cyt, respectively. A side view (Fig. 3B) of the DNA duplex shows nucleotides (D4Cyt through D7Cyt) from chain D interacting with the loop of the HhH2 unit. Salt bridges link the side chain of Lys-195 and the phosphate group of D5Thy and of Lys-198 and Arg-215 with D7Cyt (Fig. 3, A and B). Within the HhH2 loop, Arg-220 binds in the minor groove of the duplex DNA and interacts with the ribose ring oxygen atoms of both D5Thy and D6Gua and with the thymine ring carbonyl of D5Thy (Fig. 3, A and B). Interaction of the HhH2 motif and the duplex DNA probably control the spacing and alignment of the DNA substrate for enzymatic cleavage.

Interactions of the T4 RNase H with the DNA Branch Point and the 5’ Arm—Three α-helices, H1, H2, and H4, separate the two DNA strands; H1 interacts with the branch point, H2 with the 3’ arm, and H4 with the 5’ arm of the fork DNA substrate (Fig. 2C). The scissile bond is located between C7Thy and C8Thy within the DNA duplex (Fig. 1B). At the C terminus of H1, D12Ade and C7Thy make hydrophobic contacts with Val-30 and Leu-29, respectively, and the branch point base pair tilts slightly (Fig. 3C). Abutment of Leu-29 against C7Thy and the interaction of His-174 with the base pair of C8Thy:D11Ade in the major groove help position the scissile bond over Asp-155, a conserved active site residue (Fig. 3C). A stereo image of the electron density of the crucial interactions at the branch point is presented in Fig. 3D.

Helix H4 is located immediately after the bridge region, a disordered segment (residues 91–96) above the active site (Fig. 2C). Residues Trp-101 and Phe-105 of H4 form a π-stack with the first and second unpaired bases of the 5’ arm, C6Thy and C5Cyt (Fig. 3, C and D). Base-stacking occurs within the single-stranded 5’ arm, between C4Ade and C5Cyt (Fig. 3C) and between C3Ade and C2Thy (not shown). His-109 contacts C5Cyt perpendicular to the ring (cation–π) and makes an H-bond with the ring nitrogen atom of C4Ade. C2Thy is in contact with the blunt end of the DNA duplex of an adjacent symmetry molecule, and C1Cyt is disordered.

Changes in the Structure upon Substrate Binding—The structure described in the present report is a complex of a fork DNA with the T4 RNase H enzyme that has a D132N substitution in the active site. Both an inactive mutant (with a D132N mutation) and the use of the chelating agent EDTA were essential to grow these crystals. Presumably, both were required to eliminate any trace of residual activity and hydrolysis of the DNA substrate. In the protein–DNA crystal structure described here, divalent cations required for exonuclease activity were not present in the active site of the enzyme. A previous report detailed our structural work on the wild type form of the enzyme with two magnesium ions bound in the active site (PDB code 1TRF) (21). After superposition of these two structures, the root mean square deviation was found to be 1.05 Å over 283 residues: 12–88, 98–180, and 183–305 (Fig. 4A). This comparison facilitated plausible positioning of the magnesium ions adjacent to the scissile bond in the DNA–protein complex (Fig. 4B) and placement of the DNA substrate in the active site of native metal-bound RNase H (Fig. 4C). For subsequent analysis of the active site, it was assumed that the magnesium ions are in the same position in the presence of a DNA substrate as they are in the native structure. Analyses of other nucleases suggest that active site rearrangements can occur on substrate binding (43).
We are currently attempting structural studies of the T4 RNase H enzyme (native and D132N mutant) in the presence of both the fork DNA and the metal ions in the active site. Helices H7, H8, H9, and H10 comprise the small \(H_9251\)-helical subdomain and contain the \(HhH2\) motif. Two loops, H10-H11 and S5-H7, tether the small subdomain to the large subdomain; S5-H7 is linked to the \(HhH2\) motif (Fig. 4A). In the native structure, two magnesium ions were identified with Mg2 fully hydrated, and the Asp-132 inner sphere was coordinated to Mg1. The superposition of original T4 RNase H model (brown) with the new T4 D132N RNase H (gray) allows positioning of the two hydrated magnesium ions (yellow) with water molecules (red) from native T4 RNase H into the active site of the new complex. With substrate bound, the bridge region (gray spheres) remains disordered, whereas the S5-H7 loop becomes ordered. Concomitantly, the small domain rotates inward toward the duplex, and this motion brings the N-terminal portion of the \(HhH2\) loop about 4 Å closer to the S5-H7 loop. As part of this rotation, loop S2-H3 (residues 73–78), which forms a ridge on the back side of the cleft, moves slightly, thus opening the active site (Fig. 4A).

Helices H1–H6 and H11–H13 plus five parallel \(H_9252\)-strands (S1–S5) make up the large subdomain of the enzyme. The segment between helices H3 and H4 (residues 83–101) forms an open bridge above the active site. This region appears as a hel-
ical arch in the comparable structure of the 5′ to 3′ exonuclease from the bacteriophage T5 (22). In the structure of the native T4 RNase H, a segment of the bridge (residues 89–96) was too disordered to be modeled. It was initially speculated that the bridge segment would become ordered when substrate was bound, considering that the large number of basic residues in this region (Lys-87, Lys-88, Arg-90, Lys-92, and Arg-94) would be available for binding to the negatively charged DNA. With DNA in the active site, however, none of the basic residues seemed to interact with the 5′ arm of the substrate. Moreover, bridge residues 91–96 remained disordered. Polypeptide backbone segments of Met-173 and His-174 moved ~2 Å, enabling the His-174 side chain to interact with the penultimate base pair (Figs. 4D and 3C). More dramatically, a displacement of nearly 5 Å of the side chain of Gln-22 led to the formation of a hydrogen bond with Ser-153 and Ser-154, both situated next to the active site (Fig. 4D).

Residues Asn-89 and Thr-98, located immediately before and after the disordered bridge, residues Asp-8 and Ser-182, located in disordered regions of the native structure (PDB code 1TFR) and residue Lys-35 located in the bend of a loop are in the generously allowed region of the Ramachandran plot (Table 1). Other structural differences can be attributed to lattice contacts. The N-terminal 11 residues, completely disordered in the native structure (PDB code 1TFR), are modeled in this new structure adjacent to a neighboring molecule. The 3′ arm of the fork DNA is positioned nearby but does not appear to interact with the N-terminal segment. Also, the conformation of residue Lys-128 is distorted by the lattice contact between protein and the blunt end of a symmetry-related DNA molecule. This is the residue in the disallowed region of the Ramachandran plot (Table 1).

**DISCUSSION**

Most structures of FEN-1 enzymes have two divalent metal ions positioned in highly conserved active sites (21, 22, 24, 25, 28). A two-metal mechanism has been proposed for the FEN-1 family, analogous to that proposed for the 3′ to 5′ exonuclease domain of DNA polymerase I from *E. coli* (2, 44). In this mechanism, the metal ion in site A facilitates the formation of a hydroxyl anion that carries out a nucleophilic attack on the scissile phosphate. A metal ion in site B stabilizes both the oxy-anion of the leaving group and the phosphorane intermediate.

With the fork DNA as substrate, the assay of nuclease activity of T4 RNase H indicated a primary hydrolysis site at C8P between bases C7Thy and C8Thy, thereby releasing a 7-mer oligonucleotide (Fig. 1B). Secondary cleavages were detected at C6P (releasing a 5-mer) and at C9P (releasing an 8-mer), but hydrolysis at C7P was limited. With respect to the scissile phosphate, the location of Mg2 in the active site of native T4 RNase H does not favor a two-metal mechanism. Both metals in the composite structure (Fig. 4C) appear to be too far away for inner sphere coordination to the phosphates. Moreover, the two-metal mechanism requires the distance between the metal ions to be less than 4 Å. Interestingly, the crystal structures of the native T4 RNase H and the archaeal *Mj* and *Pfu* FEN-1 enzymes show that the two bound metal ions in the active site are ~7, 5, and 5 Å apart, respectively (21, 24, 25). These observations suggest that either the two magnesium ions move closer together upon DNA binding (24) or, alternatively, that FEN-1 enzymes carry out hydrolysis by a different two-metal mechanism.

Although most nucleases utilize a two-metal mechanism (43), it seems possible for T4 RNase H to manifest activity through a metal-activating process where active site residues act as a base and/or acid. In this type of nuclease activity, the role of the metal is to coordinate the substrate and shield the negative charges in the active site. In the composite structure, Asp-155 or Asp-19 could act as the base activating a water molecule for nucleophilic attack on the substrate, and Lys-199 or Lys-87 could then act as the acid, donating a proton for product release (Fig. 4B). Lys-199 interacts with Mg2 through an intervening water molecule in the native T4 RNase H (21). In mutational studies (29), Lys-87 was identified as an important contact residue for DNA binding, as evidenced by a 50% decrease in DNA binding affinity when alanine was substituted for Lys-87 in T4 RNase H. In the absence of an active site rearrangement, the general reaction mechanism for the FEN-1 family probably differs from that proposed for the Klenow fragment. Further studies will be required for a more comprehensive understanding of the mechanisms of action of this family of enzymes (24, 45–47).

**Substrate Recognition by the FEN-1 Family of Enzymes**—Several models for a 5′-nuclease-DNA complex of this family of enzyme have previously been proposed (54, 55). In Fig. 5 the structure of T4 RNase H (gray) with its branched substrate is superimposed on that of *Afu* FEN-1 (bronze) (PDB code 1RXW), the only other available structure of a FEN-1 nuclease complexed with DNA (23). The ligand in the latter complex was a DNA duplex containing a single 3′ base overhang (23). In the superposition, the 3′ fork arm from T4 structure (magenta) overlaps the corresponding strand of the duplex from the *Afu* structure (green). The 3′ flap of the *Afu* duplex (red) resides in a pocket above the C-terminal helices. Comparison of duplex arms of the two superimposed DNA substrates indicates that a flap substrate was oriented at an angle of 110–120° between the duplex axes, approximating the angle 90–100° suggested by fluorescence resonance energy transfer experiments (23) (Fig. 5A). The 3′ overhang binding site and the 5′ cleavage site are separated by ~26 Å, positioned on opposite sides of helices of the large domain of T4 RNase H (Fig. 5A).

Sequence alignment of the FEN-1 family revealed a variable region (30) appearing as a helical arch in the T5 exonuclease and as “helical clamp” structures in the FEN-1 enzymes (22, 24, 25). Presumably, the helical clamps located above the active site are involved in recognition of the 5′ arm, provided that a conformational change occurs to assist in substrate binding. Superposition of the T4 complex and *Afu* FEN-1 indicate that the 5′ arm in the T4 structure threads under the helical clamp of *Afu* FEN-1 (Fig. 5A and supplementary movie 5A associated with Fig. 5A) as predicted (48). However, the disordered “bridge” in T4 RNase H does not adopt a discernable specific structure in the presence of DNA.

Eukaryotic and archaeal FEN-1 enzymes exhibit their highest activities on double-flap DNA substrates, which have a single 3′ base overhang (39). T4 RNase H displays similar cleavage specificity but with significantly reduced activity on double-flap substrates (Fig. 1C). In the structure of *Afu* FEN-1, the sugar
moiety on the 3' overhang fits into a pocket formed by loops α2-α3 and α14-α15. The 3'-hydroxyl group is within hydrogen-bonding distance of three residues in these loops (Lys-48, Thr-55, and His-308, Fig. 5B). Superposition of the T4 and Afu nucleases shows that T4 RNase H has an additional turn of helix H2 (superimposes on Afu helix α3), a shorter loop H12-H13 (Afu loop α14-α15), and that the T4 H1-H2 loop differs in length and position from the Afu α2-α3 loop. Thus, the 3' flap pocket does not seem to be present in the T4 RNase H structure, as no residue is available for hydrogen bonding to the substrate. Moreover, T4 Ser-41 resides where the 3'-OH of the double-flap would bind, effectively eliminating selective binding of a 3' overhang (Fig. 5B). These differences in the structures of the T4 and Afu enzymes might explain why the double-flap substrate is not preferred by T4 RNase H.

Interactions between T4 RNase H, T4 45 Clamp, and T4 32 Protein on the Lagging Strand—The structure of T4 RNase H complexed with fork DNA provides important insight into how the nuclease integrates into the rest of the T4 replication system (Fig. 6). In Fig. 6A, chain D with the 3' arm (magenta) of the fork and chain C with the 5' arm (cyan) correspond to the lagging strand template and fragment, respectively. Before the polymerase has finished elongating the upstream fragment in normal replication, T4 RNase H removes the RNA primers and some adjacent DNA in a 32 protein-dependent reaction (Fig. 6B). T4 RNase H stimulation by the 45 clamp likely serves a back-up function by loading the nuclease on nicks formed when polymerase finishes the upstream fragment before the primers have been removed by the nuclease (Fig. 6C) (9, 49).

The N-terminal sequence responsible for clamp binding matches a widely conserved group adopting a common clamp interaction motif (50). This motif is found in the C-terminal region of the DNA polymerase of phage T4 and its close relative RB69, as well as near the C-terminal segment of FEN-1 nucleases. A peptide with the sequence of the clamp binding motif of the RB69 phage DNA polymerase has been shown to bind in a hydrophobic pocket on the interdomain loop of the RB69 clamp (51). The latter is analogous to the position of the clamp binding sites for the Afu and human FEN-1 proteins (23, 28). In the T4 RNase H structure, the N-terminal clamp binding motif is close to the C-terminal binding site for the 32 protein (Fig. 6D). It is also located in a position relative to the active site that is similar to the position of the C-terminal clamp binding regions in other FEN-1 family nucleases (23, 28).

A model of the ternary complex of the nuclease, DNA, and 32 protein was constructed by positioning the core of 32 protein (PDB code 1GPC) (52) adjacent to the end of the 3' fork arm in the T4 RNase H-DNA complex (Fig. 6D). The selected core structure of the 32 protein contains residues 22–239 of the 301 in the full-length protein. Residues 213–239 (blue in Fig. 6D) are part of a region (residues 213–255) thought to interact with T4 DNA polymerase (53). With the orientation of the 32 core used in Fig. 6D, the polymerase binding site is located on the opposite side from the nuclease. Polarity of the single-stranded DNA within the 32 protein remains uncertain and, therefore, was not stipulated in the model of the ternary complex. In studies of T4 RNase H mutants, deletion of C-terminal residues 286–305 (green in Fig. 6D) abolished stimulation of the nuclease by the 32 protein. Mutation of L3A or M6A or deletion of the N-terminal residues 2–10 (yellow in Fig. 6D) abolished its stimulation by the 45 clamp (9).

4 Y. Shamoo, personal communication.
C termini strongly suggests that the binding of the 32 protein would interfere with the binding of the 45 clamp and vice versa.

In our structure, three bases of the 5′ arm of fork DNA participate in specific interactions with the nuclease that are important for substrate recognition, substrate binding, and proper substrate cleavage (Fig. 3C). However, these interactions could not take place on the gapped and nicked DNA substrates produced during lagging strand replication (Fig. 6, B and C). Therefore, it seems likely that the unwinding of a small number of base pairs in the gapped and nicked DNA substrates may be necessary for proper activity of the enzyme. Consistent with this suggestion, the predominant products of hydrolysis of gapped substrates by T4 RNase H alone or in a complex with the 32 protein are dimers and trimers rather than monomers (5, 9, 49). This generalization also holds for the hydrolysis of nicked substrates by the nuclease affixed to the T4 45 clamp.

The structure of T4 RNase H in the presence of a fork DNA substrate has direct implications for the understanding of DNA recognition by the FEN-1 family of enzymes. The HhH2 motif is crucial for substrate binding of the DNA duplex, while the 5′ arm threads below the bridge/helical arch region. Three bases closest to the branch point interact with aromatic residues of the enzyme. Comparison with the FEN-1 enzyme reveals that the 3′ flap binding pocket is not present in T4 RNase H, an observation that explains its substrate preference for single rather than double flaps. This novel structure also provides important insights into the processes of lagging strand synthesis, since the 3′ arm of the DNA is situated in an optimal position for binding both the T4 32 protein and the T4 45 clamp.

Acknowledgments—We thank Dr. Allen Edmundson and Dr. Max O. Funk for critical review of the manuscript and valuable comments.

Bacteriophage T4 5′ Nuclease Bound to Branched DNA

REFERENCES

1. Henneke, G., Friedrich-Heineken, E., and Hubscher, U. (2003) Trends Biochem. Sci. 28, 384–390
2. Liu, Y., Kao, H. L., and Bambara, R. A. (2004) Annu. Rev. Biochem. 73, 589–615
3. Shen, B., Singh, P., Liu, R., Qiu, J., Zheng, L., Finger, L. D., and Alas, S. (2005) BioEssays 27, 717–729
4. Hollingsworth, H. C., and Nossal, N. G. (1991) J. Biol. Chem. 266, 1888–1897
5. Bhagwat, M., Hobbs, L. J., and Nossal, N. G. (1997) J. Biol. Chem. 272, 28523–28530
6. Hobbs, L. J., and Nossal, N. G. (1996) J. Bacteriol. 178, 6772–6777
7. Woodworth, D. L., and Kreuzer, K. N. (1996) Genetics 143, 1081–1090
8. Biswas, E. E., Zhu, F. X., and Biswas, S. B. (1997) Biochemistry 36, 5955–5962
9. Gangisetty, O., Jones, C. E., Bhagwat, M., and Nossal, N. G. (2005) J. Biol. Chem. 280, 12876–12887
10. Gomes, X. V., and Burgers, P. M. (2000) EMBO J. 19, 3811–3821
11. Li, X., Li, J., Harrington, L., Lieber, M. R., and Burgers, P. M. (1995) J. Biol. Chem. 270, 22109–22112
12. Tom, S., Henricksson, L. A., and Bambara, R. A. (2000) J. Biol. Chem. 275, 10498–10505
13. Brosh, R. M., Jr., von Kobbe, C., Sommers, J. A., Karmakar, P., Opresko, P. L., Piotrowski, J., Dianova, I., Dianov, G. L., and Bohr, V. A. (2001) EMBO J. 20, 5791–5801
14. Sharma, S., Sommers, J. A., Gary, R. K., Friedrich-Heineken, E., Hubscher, U., and Brosh, R. M. Jr., (2005) Nucleic Acids Res. 33, 6769–6781
15. Sharma, S., Sommers, J. A., Wu, L., Bohr, V. A., Hickson, I. D., and Brosh, R. M. Jr. (2004) J. Biol. Chem. 279, 9847–9856
16. Wang, W., Brandt, P., Rossi, M. L., Lindsey-Boltz, L., Podust, V., Fanning, E., Sancar, A., and Bambara, R. A. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 16762–16767
17. Ayyagari, R., Gomes, X. V., Gordenin, D. A., and Burgers, P. M. (2003) J. Biol. Chem. 278, 1618–1625
18. Garg, P., Stith, C. M., Sabourin, N., Johansson, E., and Burgers, P. M. (2004) Genes Dev. 18, 2764–2773
19. Kao, H. I., Veeraraghavan, J., Polaczek, P., Campbell, J. L., and Bambara, R. A. (2004) J. Biol. Chem. 279, 15014–15024
20. Maga, G., Villani, G., Tillement, Y., Stucki, M., Locateelli, G. A., Fouzin, I., Spadari, S., and Hubscher, U. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 14298–14303
21. Mueser, T. C., Nossal, N. G., and Hyde, C. C. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 9847–9853
22. Bosbeck, P., Mueser, T. C., Nossal, N. G., and Hyde, C. C. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 90–93
23. Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J. A. (2004) Cell 116, 39–50
24. Hosfield, D. J., Mol, C. D., Shen, B., and Tainer, J. A. (1998) Cell 95, 135–146
25. Hwang, K. Y., Chakravarthy, P., Karmakar, P., Opresko, P. L., Piotrowski, J., Dianova, I., Dianov, G. L., and Bohr, V. A. (2001) EMBO J. 20, 5791–5801
26. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995) Nature 376, 612–616
27. Matsui, E., Musti, K. V., Abe, J., Yamashita, K., Matsu, I., and Harata, K. (2002) J. Biol. Chem. 277, 37840–37847

FIGURE 6. Interactions of T4 RNase H with the T4 32 single-stranded DNA-binding protein and the T4 gene 45 clamp on the lagging strand. A, T4 RNase H bound to the fork DNA in the structure. The 3′ arm strand (magenta) corresponds to the lagging strand template in panels B and C, whereas the 5′ arm strand (cyan) corresponds to the lagging strand Okazaki fragment. B, T4 RNase H normally removes the RNA primer and some adjacent DNA when stimulated by the 32 protein bound behind it (9). Simultaneously, the progressing polymerase-clamp complex completes the previous fragment and displaces the 32 protein. C, if polymerase completes the previous fragment before RNase H removes the primer, a nicked duplex is formed. The T4 45 clamp increases the rate of hydrolysis of nicked duplexes by T4 RNase H. D, the core of the T4 32 protein is docked with the T4 RNase H-DNA complex. This T4 32 core (PDB code 1GPC, orange) (52) docked close to the C-terminal residues of RNase H (286–305, green) that are required for its interaction with the full-length 32 protein. The C-terminal segment (residues 213–239, blue) of the 32 core structure is part of a region (residues 213–255) believed to interact with T4 DNA polymerase (53). The N-terminal residues of T4 RNase H (1–11) required for stimulation of RNase H by the 45 clamp are shown in yellow. See “Discussion” for a further discussion of this model.
Bacteriophage T4 5’ Nuclease Bound to Branched DNA

28. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., and Hakoshima, T. (2005) EMBO J. 24, 683–693
29. Bhagwat, M., Meara, D., and Nossal, N. G. (1997) J. Biol. Chem. 272, 28531–28538
30. Harrington, J. I., and Lieber, M. R. (1994) EMBO J. 13, 1235–1246
31. Jones, C. E., Mueser, T. C., and Nossal, N. G. (2004) J. Biol. Chem. 279, 12067–12075
32. Matthews, B. W. (1968) J. Mol. Biol. 33, 491–497
33. Navaza, J., and Saludjian, P. (1997) Methods Enzymol. 276, 581–594
34. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. D Biol. Crystallogr. 54, 905–921
36. Murshudov, G. N. (1997) Acta Crystallogr. D Biol. Crystallogr. 53, 240–255
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1992) J. Appl. Crystallogr. 26, 283–291
38. Krissinel, E., and Henrick, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268
39. Kaiser, M. W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L., and Lyamichev, V. I. (1999) J. Biol. Chem. 274, 21387–21394
40. Qiu, J., Liu, R., Chapados, B. R., Sherman, M., Tainer, J. A., and Shen, B. (2004) J. Biol. Chem. 279, 24394–24402
41. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) Nucleic Acids Res. 24, 2488–2497
42. Shao, X., and Grishin, N. V. (2000) Nucleic Acids Res. 28, 2643–2650
43. Yang, W., Lee, J. Y., and Nowotny, M. (2006) Mol. Cell 22, 5–13
44. Beece, L. S., and Steitz, T. A. (1991) EMBO J. 10, 25–33
45. Tock, M. R., Frary, E., Sayers, J. R., and Grasby, I. A. (2003) EMBO J. 22, 995–1004
46. Feng, M., Patel, D., Dervan, J. P., Ceska, T., Suck, D., Haq, I., and Sayers, J. R. (2004) Nat. Struct. Mol. Biol. 11, 450–456
47. Cowan, J. A. (2002) Biometals 15, 225–235
48. Murante, R. S., Rust, L., and Bambara, R. A. (1995) J. Biol. Chem. 270, 30377–30383
49. Bhagwat, M., and Nossal, N. G. (2001) J. Biol. Chem. 276, 28516–28524
50. Warbrick, E. (1998) BioEssays 20, 195–199
51. Shamoo, Y., and Steitz, T. A. (1999) Cell 99, 155–166
52. Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H., and Steitz, T. A. (1995) Nature 376, 362–366
53. Sun, S., and Shamoo, Y. (2003) J. Biol. Chem. 278, 3876–3881
54. Dervan, J. P., Feng, M., Patel, D., Grasby, J. A., Artyymiuk, P. J., Ceska, T. A., and Sayers, J. R. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 8542–8547
55. Allawi, H. T., Kaiser, M. W., Onufriev, A. V., Ma, W. P., Brogaard, A. E., Case, D. A., Neri, B. P., and Lyamichev, V. I. (2003) J. Mol. Biol. 328, 537–554