Molecular Structure and Novel DNA Binding Sites Located in Loops of Flap Endonuclease-1 from Pyrococcus horikoshii*

Eriko Matsui†, Krishnasasstry V. Musti‡§, Junko Abe†, Kazuhiko Yamasaki†, Ikuo Matsui‡,§, and Kazuaki Harata‡

From the †Biological Information Research Center and the ‡Gene Discovery Research Center, National Institute of Advanced Industrial Science and Technology, Higashi 1-1-1, Tsukuba, Ibaraki 305-656, Japan

The crystal structure of flap endonuclease-1 from Pyrococcus horikoshii (phFEN-1) was determined to a resolution of 3.1 Å. The active cleft of the phFEN-1 molecule is formed with one large loop and four small loops. We examined the function of the conserved residues and positively charged clusters on these loops by kinetic analysis with 45 different mutants. Arg40 and Arg42 on small loop 1, a cluster Lys193–Lys195 on small loop 2, and two sites, Arg94 and Arg118-Lys119, on the large loop were identified as binding sites. Lys87 on the large loop may play significant roles in catalytic reaction. Furthermore, we successfully elucidated the function of the four DNA binding sites that form productive ES complexes specific for each endo- or exo-type hydrolysis, probably by bending the substrates. For the endo-activity, Arg34 and Lys183–Lys186 located at the top and bottom of the molecule were key determinants. For the exo-activity, all four sites were needed, but Arg118-Lys119 was dominant. The major binding sites for both the nick substrate and double-stranded DNA might be the same.

A hyperthermophilic archaeon Pyrococcus horikoshii was isolated from a hydrothermal vent in the Okinawa trough in the Pacific Ocean (1). It grows optimally at around 100 °C. The entire genome sequence of this archaen has been determined by the National Institute of Technology and Evaluation in Japan (2). Our current understanding of the DNA replication mechanism in archaea, the third domain of life, is still rudimentary. However, several genes encoding eukaryotic-like DNA replication proteins are present in archaea genomes (2–5). This observation has led to the proposal that the archaea and eukaryotic DNA replication mechanisms share important similarities.

Flap endonuclease-1 (FEN-1) has important roles in DNA replication, repair, and recombination. FEN-1 has 5′-endonuclease and 5′-3′-exonuclease activities. In DNA replication, FEN-1 removes the RNA primers during the maturation of the Okazaki fragment (6–9). For DNA repair, FEN-1 removes damaged nucleotides after apurinic/apyrimidinic endonuclease has incised the 5′ side of the apurinic/apyrimidinic site in long patch base excision repair (10–12). FEN-1 is also required for nonhomologous DNA end joining of double strand DNA breaks (13). The FEN-1 sequence is conserved among eukaryotes and archaea (14–17). Two crystal structures of FEN-1 have been reported exclusively in thermophilic archaea (18, 19). The molecular structures of the members of the FEN-1 family, T5 exonuclease, T4 RNase H, and the exonuclease domain of Taq polymerase, were also reported (20–22). They have in common a large helical arch mounted upon a globular domain containing the active site. It was postulated that the flap strand of the substrate DNA threads through this arch (22). Several studies have provided evidence of a role for the arch in tracking the flap strand (23–25). However, it remains unclear whether the flap strand of the substrate DNA threads through this arch because FEN-1 could cleave the flap strand with a secondary structure even at a reduced rate and could efficiently hydrolyze the branched structure (26, 27).

The FEN-1 homologue (PH1415) was identified in the P. horikoshii genome (2) and then overexpressed in Escherichia coli, and the recombinant protein was characterized in detail using 35 different substrates (17). FEN-1 possesses 5′-flap endonuclease and 5′-3′-exonuclease activities. The flap endonuclease cleaves the flap strand at the junction, and the activity is independent of the length of the 5′ flap strand, cleaving both the 1- and 5-mer flap strands efficiently as well as the 19-mer flap strand (28). The secondary structure of flap strand inhibited the endonuclease activity of eukaryotic FEN-1 (25–27). Archaeal FEN-1 could cleave the double flap strand (15, 17). The flap endonuclease activity has been shown to require the upstream primer, which fills up until the junction portion, and expanding the 3′ of the upstream primer elevates the activity (16, 17, 23). The 5′-3′-exonuclease activity digested the double strand DNA containing the nick, gap, and 5′ recess-end, and the activity was elevated by expanding the 3′ end of the upstream primer in the nick portion (28, 29). There have been a few studies concerning the dual function of acidic residues located at the active center, which binds to both the substrate and active Mg2+ ions (30–32). Site-directed mutagenesis of T5 5′-3′-exonuclease showed that a conserved lysine residue (Lys85) located near the active site was required for exonuclease activity but not for endonuclease activity (33). This finding indicates a difference in substrate recognition for exo- and endonuclease activities.

The substrate specificity of FEN-1 was investigated in detail...
as described above, whereas the DNA-binding mechanism, including the DNA-binding sites and the function of the large loop, is still unclear. To investigate the substrate-binding mechanism of phFEN-1, we determined the molecular structure of phFEN-1 as reported here. On the basis of this structure, we analyzed the function of one large and four small loops by site-directed mutagenesis, and identified several key amino acid residues involved in endo- and exo-type DNA binding and catalysis.

**EXPERIMENTAL PROCEDURES**

**Chemicals**—The vector pET-11b and *E. coli* strain BL21 (DE3) were purchased from Novagen (Madison, WI). The vector pGEMEX was obtained from Promega (Madison, WI). The restriction enzymes were purchased from Promega. Isopropyl-β-D-thiogalactopyranoside was also obtained from Takara Shuzo. The oligonucleotides labeled with fluorescent dye, fluorescein (FAM), were purchased from Sawaiy Technology, Ltd. (Tokyo, Japan).

**Expression, Purification of Recombinant phFEN-1, and Measurement of CD**—Each mutant gene was constructed on the basis of the vector pGEMEX/phFEN-1 using the site-directed mutagenesis procedure and the PCR method. The wild-type and mutated genes were inserted into the vector pGEMEX using NdeI and XhoI sites, resulting in the expression vectors pET-11a/phFEN-1s.

The *E. coli* strain BL21 (DE3) was transformed with the expression vectors. The transformant was propagated in 200 ml of 2X yeast extract medium containing ampicillin at 37 °C. The recombinant protein was induced at A_{600} = 1.0 − 1.5 with 1 mM of isopropyl-1-thio-β-D-galactopyranoside for 12 h. The cells were collected by centrifugation and stored at −20 °C. The frozen cells were then thawed at room temperature and mixed with 1 ml of 50 mM Tris-HCl buffer (pH 8.0). The sample was heated at 83 °C for 30 min and centrifuged at 2300 × g for 30 min, and then the supernatant was again heated at 83 °C for 15 min to remove endogenous proteases completely before being centrifuged. The supernatant was loaded on a Hitrap SP column (Amersham Biosciences) and eluted with a linear gradient of 1 M NaCl in a 50-mM Tris-HCl buffer containing 20% PEG 6000 with an equal amount of reservoir solution containing 20% PEG 6000 and 2% protein and 200 mM NaCl, pH 8.0 (50 mM Tris-HCl buffer). The recombinant phFEN-1s were eluted with buffer containing 200−300 mM NaCl.

The secondary structure of the purified proteins was measured using a circular dichroism (CD) spectrometer, AVIV model 62DS (Shimazu, Kyoto, Japan) and a quartz cuvette having a path length of 1.0 or 2.0 mm. The protein concentration was −0.1 mg/ml of 50 mM Tris-HCl (pH 8.0).

**Structural Determination**—Crystals were obtained by hanging-drop vapor diffusion using the R42E mutant protein. To a 2-μl solution containing 2% protein and 200 mM NaCl, pH 8.0 (50 mM Tris-HCl buffer), an equal amount of reservoir solution containing 20% PEG 6000 and 1 M NaCl was added to the drop solution, and the cell dimensions were a = b = 62.67 Å and c = 150.69 Å. The asymmetric unit contained two independent molecules (V_{as} = 2.56 Å³/Da). The intensity data were measured to a resolution of 3.0 Å using synchrotron radiation (λ = 1.0 Å) on an ADSC Quantum 4R charge-coupled devices detector at the BL-6A beam line station of the Photon Factory. A total of 135,478 observed reflections were merged to a set of 15,903 (100% completeness) with an R_mer of 0.104.

The structure was solved by the molecular replacement method using the structure of the FEN-1 from *Methanococcus jannaschii* (37). A self-rotation search using the program X-POLR indicated that the two independent molecules are related by a local 2-fold axis. The structure was refined by X-FLOR for 13,576 reflections in the resolution range from 0.0 to 3.1 Å. The reflection converged at R and R_{free}, respectively, of 0.190 and 0.279, respectively. The root mean square deviations of bond distances and angles from their ideal values were 0.020 Å and 3.1 °, respectively. The coordinates were deposited in the Protein Data Bank (ID code 1MCB).

**Oligonucleotides Used and Preparation of Substrates**—The following oligonucleotides were designed to construct various DNA substrates.

- Oligonucleotide A, 5′-GGCTAGATGTCGAGCTTCGCTCAT-GACCGTATGCAACCTGTCGACTCTGGA-3′.
- Oligonucleotide B, 27-mer, 5′-TGGACCTGGTATGCACTCTGCTCCATGGC-3′.
- Oligonucleotide C, 57-mer, 5′-GCATCGGCTAGGCGTCGAGCTTGACACTCTGCTCCATGGC-3′.
- Oligonucleotide D, 27-mer, 5′-CTTGGACAGCTGTCGACTCTGCTCCATGGC-3′.
- Oligonucleotide E, 54-mer, 5′-TGGACTGGTATGCACTCTGCTCCATGGC-3′.
- Oligonucleotide F, 37841-mer, 5′-GCAGAGTCCGACATCTAGCTC-3′.
- Oligonucleotide G, 37841-mer, 5′-GCAGAGTCCGACATCTAGCTC-3′.

Oligonucleotides C and D were labeled at the 5′ terminus with a fluorescence group (FAM) and then purified using high pressure liquid chromatography. Annealing reactions to construct various DNA substrates were performed as described previously (38). The flap substrate was constructed with oligonucleotides A, B, and C labeled with FAM. The nick substrate was constructed with oligonucleotides A, B, and D labeled with FAM. The dsDNA was constructed with oligonucleotides A and E.

**Kinetic Analysis of Mutants Using Substrates with Fluorescence**—The levels of activity of phFEN-1s were measured using substrates with a highly fluorescent dye, fluorescein (FAM), at the 5′ end of the flap substrate or 5′ end of the unprimer of the nick substrate. The pattern observed for the digestion of fluorescent substrate was the same as that observed for the radiolabeled substrate. The product amount was evaluated with a fluorescence scanner after electrophoresis. This method has almost the same sensitivity as the radioactive-labeling methods employed for the phFEN-1 assay. Hence, we used the fluorescent substrate throughout the study.

**Measurement of Kinetic Parameters**—The kinetic parameters were determined using the fluorescence substrate. The reaction mixture was 50 mM Tris-HCl (pH 7.4) containing the labeled substrate, 1.5 mM MgCl₂, and 1.5 μg of bovine serum albumin. The enzyme and substrate concentrations were varied from 0.025 to 200 ng and from 0.1 to 75 pmol, respectively. A 10-fold volume (150 μl) of the reaction mixture was incubated at 60 °C for 15 min. The total volume of the stop solution was 100 μl as a control for a zero time reaction. Nine volumes of enzyme solution were added, and 15 μl of the reaction mixture was added to 10 μl of the stop solution every 1 min for 7 min. The samples were denatured at 95 °C for 10 min. Five μl of sample was loaded onto a 12% polyacrylamide gel (10 × 10 cm) with 7 μl urea and 1X TBE, and electrophoresed with the same buffer for 1.5 h at 200 V. The reaction products were visualized and quantified using a FluorImager 585 (Molecular Dynamics, Inc., CA). The initial velocities were obtained directly from the initial slopes of the time course plots. The K_{m} and k_cat values were calculated using the Michaelis-Menten equation and the least squares method (34).

**Surface Plasmon Resonance Experiments**—The interaction between phFEN-1 and dsDNA was determined using a highly fluorescent dye, fluorescein (FAM), at the 5′ end of the flap substrate or 5′ end of the unprimer of the nick substrate. The pattern observed for the digestion of fluorescent substrate was the same as that observed for the radiolabeled substrate. The product amount was evaluated with a fluorescence scanner after electrophoresis. This method has almost the same sensitivity as the radioactive-labeling methods employed for the phFEN-1 assay. Hence, we used the fluorescent substrate throughout the study.

**RESULTS**

**Three-dimensional Structure of phFEN-1**—The crystallization was investigated with the wild type and some mutant enzymes, and a rod-shaped crystal was obtained with the R42E mutant. The crystal structure was solved by molecular replacement using the structure of FEN-1 from *M. jannaschii* and was refined at a resolution of 3.1 Å. The two independent molecules are related by the local 2-fold axis as shown in Fig. 1A. The structure consists of 10 helices, A-J; five β-strands, A-E; and loop regions. The helices A, B, C, I, and J are separated from the other five helices by the sheet-like structure of the five β-strands. The arrangement of the helices and β-strands is similar to that observed in the FEN-1 structures of *M. jannaschii* and *Pyrococcus furiosus*. The region of residues 80-128 between sheet B and helix 1 forms a large and flexible loop designated the large loop. The region 187-206 between sheet E (in the ε-helix) and sheet F (in the β-helix loop) designated the large loop. The region 187-206 between sheet E (in the ε-helix) and sheet F (in the β-helix loop) designated the large loop. The region 187-206 between sheet E (in the ε-helix) and sheet F (in the β-helix loop) designated the large loop. The region 187-206 between sheet E (in the ε-helix) and sheet F (in the β-helix loop) designated the large loop. The region 187-206 between sheet E (in the ε-helix) and sheet F (in the β-helix loop) designated the large loop.
leucine at position 196 (17), the folding structure of small loop 2 differs markedly between *P. horikoshii* and *P. furiosus* as shown in Fig. 1B. The structural flexibility of these two loop regions of phFEN-1 is also indicated by the superimposition of two independent subunit molecules from the dimer structure (data not shown).

The two independent subunits form a dimer structure having a pseudo 2-fold symmetry as shown in Fig. 1A. The concave sides of the molecules are facing each other, which produces a large cylindrical cavity. The molecular contacts are observed among small loops 1 and 3 and a part of the large loop as shown in Fig. 1C. The regions 43–45 and 109–115 of one subunit are in contact with region 243–252 of the other subunit. The carboxyl group of Asp 43 has polar contacts with Lys 249 and Glu 252, but the mutated residue Glu 42 has no direct intermolecular contact. The region 109–115 has the hydrophobic residues Leu 109, Ala 110, and Leu 114 as shown in Fig. 1C. These residues are in van der Waals contacts with the main chain peptide groups of region 243–246 of the other subunit.

**Purification of Mutant Protein and CD Spectrum**—One large loop and four small loops (80–128, 39–55, 187–206, 234–249, and 257–263) were identified in the phFEN-1 molecule, as shown in Figs. 1 and 2. Since several basic amino acids and their clusters were observed on these loops that were expected to be important for binding DNA, 45 mutant genes were constructed by site-directed mutagenesis to analyze the function of these positively charged clusters. The mutation points are summarized in Fig. 2. The expression level of each mutant in *E. coli* was ~3-fold lower than that of the wild type. These mutants were purified using the same procedure as was used for the wild type. The molecular weight of each mutant was the same as for the wild type except R40E, which had a molecular weight half of that of the wild type, presumably due to proteolytic digestion. The R40E protein was not used in the subsequent experiment.

The CD profiles of the mutant proteins were measured to confirm that no significant changes had occurred by site-directed mutagenesis. As shown in Fig. 3, the CD spectra of the mutants were the same as the spectrum of the wild type. Hence, the following kinetic analysis was performed using these mutant proteins.

**Kinetic Parameters of Mutants on Four Small Loops and One Large Loop Using Flap Substrate**—As shown in Fig. 2, two pairs of basic amino acids (Arg 40 and Arg 42, and Lys 51 and Arg 53) were observed typically on both edges of small loop 1 (39–55) with highly conserved amino acids (Leu 47, Gly 44, and Gly 55) between the eukaryotes and archaea. Various mutations were introduced into this typical region to investigate the function of these residues. As shown in Fig. 4, the *K_m* of the single mutants R40G and R42G both increased 7-fold, compared with that of the wild-type enzymes (WT). The *K_m* values of R42E were 19-fold higher and 25-fold lower than those of WT, respectively. These results indicate that the mutational effect was increased in the following order: no side chain < neutral side chain < negative side chain. The tendency was magnified for the double mutants. The *K_m* value of R40G/R42G increased 10-fold, whereas that of R42Q was almost the same as the value for WT. The *K_m* and *k_cat/K_m* values of R42E were 19-fold higher and 25-fold lower than those of WT, respectively. These results indicate that the mutational effect was increased in the following order: no side chain < neutral side chain < negative side chain. The tendency was magnified for the double mutants. The *K_m* value of R40G/R42G was elevated 26-fold compared with that of the wild type, as well as R40Q/R42Q. The *K_m* of R40E/R42E was increased 105-fold compared with WT, suggesting strong repulsion among these residues and DNA. The *k_cat* and *K_cat/K_m* values

---

**Fig. 1.** The steric structure of a dimer phFEN-1 molecule. A, stereo view of the dimer. Red coils show α helices (A–J) and blue arrows show β strands (A–E). The figure was produced using the program MolScript. B, stereo view of the superimposition of *P. horikoshii* (red), *P. furiosus* (green), and *M. jannaschii* (blue) FEN-1s. The coloration of the figures was produced using the program Turbo-Frodo. C, stereo view of the dimer interface. The main chains from different subunits are distinguished with red and blue colors.

**Fig. 2.** The location of the mutation on the structure. Red coils show α helices (A–J) and blue arrows show β strands (A–E). The number indicates the position of the mutation. Small loops 1 (39–55) and 2 (187–206), and the large loop (80–128) are colored yellow, green, and pink, respectively.

---

**FIG.1.** The steric structure of a dimer phFEN-1 molecule. A, stereo view of the dimer. Red coils show α helices (A–J) and blue arrows show β strands (A–E). The figure was produced using the program MolScript. B, stereo view of the superimposition of *P. horikoshii* (red), *P. furiosus* (green), and *M. jannaschii* (blue) FEN-1s. The coloration of the figures was produced using the program Turbo-Frodo. C, stereo view of the dimer interface. The main chains from different subunits are distinguished with red and blue colors.

**FIG.2.** The location of the mutation on the structure. Red coils show α helices (A–J) and blue arrows show β strands (A–E). The number indicates the position of the mutation. Small loops 1 (39–55) and 2 (187–206), and the large loop (80–128) are colored yellow, green, and pink, respectively.
were also decreased 4- and 680-fold, respectively. On the other hand, the mutations of Lys<sup>51</sup> and Arg<sup>53</sup> showed little difference in the \( K_m \), \( k_{cat} \), and \( k_{cat}/K_m \) values compared with those of WT. Even the double mutant K51E/R53E retained 30% of the WT \( k_{cat}/K_m \) value.

The \( K_m \) value of L47G was increased 20-fold, whereas the values of a conservative mutant L47F was similar to that of WT. The \( k_{cat} \) values of both L47G and L47F were almost the same as that of WT. These findings indicated important hydrophobic interactions between Leu<sup>47</sup> and other hydrophobic side chains to maintain the typical loop structure. The mutations at Gly<sup>44</sup> and Gly<sup>52</sup> had no significant effect on the \( K_m \), \( k_{cat} \), and \( k_{cat}/K_m \) values.

Concerning small loop 2 (187–206), the basic cluster, Lys<sup>193</sup>, Arg<sup>194</sup>, and Lys<sup>195</sup>, is present on one side of the loop in which the conserved Lys<sup>199</sup> is located in the middle as shown in Fig. 2. The \( K_m \) value of each of the single mutants K193A, R194A, and K195A was 4-, 5-, and 8-fold of WT, respectively, whereas the \( k_{cat} \) values of these mutants were almost the same as for WT as shown in Fig. 4. For the triple mutant K193A/R194A/K195A, the \( K_m \) value increased markedly and the \( k_{cat} \) value decreased moderately. The negatively charged triple mutant K193E/R194E/K195E showed similar but more magnified effects on both parameters compared with the alanine triple mutant K193A/R194A/K195A, indicating ionic interactions of the positive cluster with DNA phosphate groups in an analogous manner to Arg<sup>40</sup> and Arg<sup>42</sup> on small loop 1. The values of K199A differed little from those of WT.

One DNA binding motif was putatively identified at small loop 3 (234–249) in the molecular structure of FEN-1 from P. furiosus (19). Small loop 4 (257–263) is also close to small loop 3. The positive residues, Lys<sup>243</sup>, Lys<sup>247</sup>, Lys<sup>249</sup>, and Lys<sup>263</sup>, and the conserved residue Tyr<sup>237</sup> are located inside or close to the motif as shown in Fig. 2. Several mutants were made to investigate the function of these positive or conserved residues. However, the \( K_m \) and \( k_{cat}/K_m \) values of the negative mutants (K243E and K263E) and the alanine mutants (K243A, K248A, K249A, K263A, and Y237A) did not change markedly compared with those of WT, indicating less significant roles for these positive residues compared with those on small loops 1 and 2.

The large loop (80–128) is 49 amino acids in length, and

Fig. 3. CD profiles of WT and mutant proteins. The molar ellipticity was expressed as deg cm<sup>2</sup> dmol<sup>−1</sup>. The protein concentration of each sample was ~0.1 mg/ml in 50 mM Tris-HCl (pH 8.0).

Fig. 4. Comparison of the \( K_m \), \( k_{cat} \), and \( k_{cat}/K_m \) values between WT and the mutants using flap substrate. The assay conditions were described under “Experimental Procedures.” The protein amounts for WT and the mutants were 0.025–10 ng, and the substrate concentrations were 2 nM–2 μM.
The effects of the mutations K93A and R95A on both parame-
ters were minor. The Km values were 50 nM-5 1- for WT and the mutants were 1 "cat values of R88A and K89A did not change markedly, compared
with those of WT. Interestingly, the values of R40G/R42G, R118A/K119A, and K193A/R194A/K195A decreased 222-, 1851-, and 76-fold, respectively compared with those of WT, indicating a large contribution by these two regions to the exo-type binding as shown in Fig. 5. The Km value of the single mutant R94A was decreased 200-fold compared with that of WT. The Km value of the double mutant R118A/K119A was decreased 111-fold. The kcat/Km values of R40G/R42G, R118A/K119A, and K193A/R194A/K195A decreased 222-, 1851-, and 76-fold, respectively compared with the value of WT, indicating that small loops 1 and 2 and the two regions of the large loop play important roles in recognizing the nick substrate exonucleolytically.

The single mutant K87A showed weak activity on the nick substrate, too weak to determine the kinetic parameters, suggesting the functional importance of the residue in exonucleolyis as well as endonucleolysis.

**TABLE I**

| Proteins            | $K_m$ (10^6 M⁻¹) |
|---------------------|------------------|
| WT                  | 4.14             |
| R40G                | 2.89             |
| R42G                | 2.17             |
| R40G/R42G           | 0.661            |
| K87A                | 4.2              |
| R94A                | 3.99             |
| R118A               | 2.83             |
| K119A               | 2.12             |
| R118A/K119A         | 1.13             |
| K193A/R194A/K195A   | 3.29             |

**DISCUSSION**

The positive residues in small loops 1 and 2 and the large loop contribute greatly to the flap substrate binding—Recently, two molecular structures of FEN-1s were reported from phFEN-1 could not enter the native gel. The substrate binding affinity was therefore analyzed by surface plasmon resonance measurements. A 54-mer dsDNA was immobilized on the Sensor Chip. The affinity constants ($K_a$) of WT and mutant protein are shown in Table I. The $K_a$ value of R40G/R42G and R118A/K119A dropped 6- and 4-fold, respectively, and the values for each single mutant decreased moderately compared with that of WT, whereas the $K_a$ value of the other mutants showed little difference to that of WT. The results indicated that both these sites, Arg⁴⁰ and Arg⁴², contribute to the binding of dsDNA. Furthermore, we tried to analyze the binding affinity with single-stranded DNA. The $K_a$ values of all these mutants for single-stranded DNA did not decrease significantly compared with the value for WT (data not shown).

Recently, two molecular structures of FEN-1s were reported from...
archaea (14, 18) for which differences in substrate binding were proposed. According to the two-modeled structures of the protein-substrate complex, the 5' end of the single-stranded flap strand threads through the hole formed by the L1 loop (corresponding to the large loop in our work). However, the orientation of the main double-helical strands shown in the proposed complex crosses at a 90° angle. Despite the abundance of structural information on thermostable FEN-1s from archaea, the substrate binding models were too different to comprehend the general type of binding, and little was known about how the substrates were recognized by FEN-1s endo/exonucleolytically.

To understand the substrate recognition mechanism of phFEN-1 at the molecular level, the crystal structure was determined to a resolution of 3.1 Å. phFEN-1 has a novel dimer structure in which one large loop and four small loops form a large cylindrical active cleft as shown in Fig. 1A. On the basis of the structural information for phFEN-1, 24 amino acid residues in several loops were mutated systematically to search for recognition sites in the phFEN-1 molecule (Fig. 2). The mutant proteins were purified completely, and the structural integrity was checked by a CD measurement as shown in Fig. 3. The nuclease activities were measured with 32P-labeled substrates, but the reproducibility of the reactions was too poor to allow determination of the kinetic parameters because the labeling efficiency with T4 kinase and unstable 32P was variable. Finally, a new assay method using fluorescence substrates and a FluorImager was developed, and the kinetic parameters were determined for 45 mutants, including single, double, and triple mutants at different loop regions as shown in Figs. 4 and 5.

For small loop 1 (39–55), the kinetic parameters summarized in Fig. 4 clearly indicated that the former pair, Arg40 and Arg42, bound equally to the substrate, and the latter pair, Lys81 and Arg83, played a less significant role for the substrate binding and the catalytic reaction. In the molecular structure, Leu47 was located in the middle of the two positive pairs on small loop 1 and showed tight hydrophobic interactions with the methyl and methylene groups of Ile39, Gln41 in the same loop, and Ser59 located near the loop. The kinetic parameters of Leu47 mutants, as shown in Fig. 4, strongly suggested that their hydrophobic interactions might be critical in maintaining the typical loop structure and, consequently, in adjusting the orientation of the positive side chains of Arg40 and Arg42 to the substrate. The mutations for removing the hydrophobicity of Leu47 might increase the Km value by eliminating the function of Arg40 and Arg42.

For small loop 2 (187–206), the three positive residues Lys193, Arg194, and Lys195 bind cooperatively to the DNA by ionic interaction in the same manner as Arg40 and Arg42 on loop 1 as shown in Fig. 4. However, Lys195 has a less significant function in binding the substrate than does the cluster. In the sequence alignment of small loop 2 among six FEN-1s from eukaryotes and archaea, the small loop 2 of phFEN-1 is longest, having nine extra residues, compared with those of the eukaryotic FEN-1s, as previously reported (17). As shown in Fig. 1B, phFEN-1 and pfFEN-1 show a large difference in small loop 2 folding despite their high sequence identity (95%), suggesting a marked flexibility of the loop. The conformational change of the small loop might be induced to fit the positive cluster (Lys193, Arg194, and Lys195) to bind DNA strands because it was reported that the helicity of the human FEN-1 structure was increased up to 4% by addition of DNA substrates (35).

The positive residues Lys243, Lys248, Lys249, and Lys263 are also located in small loop 3 (234–249) and small loop 4 (257–263) as shown in Fig. 2. The kinetic parameters of various mutants for these residues, even the negatively charged mutants, as shown in Fig. 4, were not changed markedly suggesting less significant roles for these positive residues compared with those in small loops 1 and 2.

The three positive clusters (87–89, 93–95, and 118–119) were located at the large loop (80–128). For the first cluster (87–89), only residue Lys87 showed drastic changes in the kcat value for endo-type hydrolysis on replacement of alanine compared with WT, although the Km value of the mutants did not change markedly as shown in Fig. 4. Our findings indicated a significant contribution of Lys87 to the catalytic mechanism. Furthermore, the functional importance of Lys87 was noteworthy, regarding the molecular structure. The Lys87 residue forms a hydrogen bond with Pro44 with a distance of 2.9 Å. Interestingly, two successive proline residues (corresponding to Pro44 and the neighboring Pro83 in the phFEN-1 molecule) are conserved completely in the initial part of the large loop (80–128) for all FEN-1s reported from eukaryotes and archaea, suggesting their structural importance (17). Because the proline residue has a fixed peptide bond angle, the two successive Pro residues, Pro43 and Pro44, might be essential in maintaining the specific structure of the initial part containing the catalytic Asp80. The mutation K87A that loses the hydrogen bond with Pro44 might cause a positional shift of the catalytic Asp40 from the active center due to a slight modification in the loop structure leading to a decreased kcat value.

In the second positive cluster (93–95), only Arg94 showed marked changes in the kinetic parameters on replacement of alanine as shown in Fig. 4. The Km and kcat values of R94A were greatly increased and decreased, respectively, clearly suggesting a markedly decreased ratio of the productive complex per total ES complex according to the subsite theory (36). Arg94 is essential to bind the substrate, whereas Lys93 and Arg95 play supplemental roles. The positive charges of Lys93 and Arg95 might be important in adjusting environmental conditions for Arg94.

In the third cluster (118 and 119), the Km value of the double mutant R118A/K119A was increased 10-fold as shown in Fig. 4, although the single mutant, R118A or K119A, showed a moderate increase in the value probably due to the functional complementarity between the two residues. These findings clearly indicate that the substrate DNA is recognized by ionic interactions with Arg94, Arg118, and Lys119 located at both typical sites of the large loop as shown in Fig. 2, whereas Lys87 located in the early part of the loop is deeply involved in the catalytic mechanism.

Substrate Binding Sites Commonly Used Exo- and Endonucleolytically—For the endo-type reaction as mentioned above, four substrate binding sites (ARG40 and ARG42, Arg94, Lys118, and Lys195) were identified and were located on small loops 1 and 2, and on the large loop, independently. FEN-1 possesses dual activities; exo-activity on dsDNA with nick and recess end, and endo-activity on the flap substrate. The same catalytic residues binding to an essential metal ion Mg2+ are indispensable for both activities. Furthermore, Garforth et al. reported an interesting mutant of T5 5'-3'-exonuclease belonging to the same family as FEN-1, which lost the exo-activity but retained the endo-activity on replacement of the Lys residue close to the active center (33). However, it is still unclear how FEN-1 recognizes the different substrates exo/endonucleolytically.

To clarify the endo- and exo-type recognition mechanism of FEN-1, kinetic parameters of exo-activity against the nick substrate were determined and are shown in Fig. 5. Using the kinetic parameters for the exo/endonucleolysis summarized in Figs. 4 and 5, the DNA binding manner specific for each activity was evaluated according to the subsite theory using following equations (36–40),
where $\Sigma K_i$ corresponds to the sum of the binding constants for all binding forms of the substrate. $K_j$ equals the sum of $K_p$ (the binding constant of ES complex leading to the product) and $K_q$ (the binding constant of ES complex failing in production). $k_{cat}$ is the intrinsic rate constant of the hydrolysis. The $k_{cat}/K_m$ value was postulated as invariant among mutant and WT enzymes, because no significant structural change had been detected by the CD measurement as shown in Fig. 3. The $k_{cat}$ of the WT enzyme for the flap substrate (1.2 s$^{-1}$) was close to the value for the nick substrate (0.7 s$^{-1}$) as shown in Figs. 4 and 5, suggesting that the same active center was used with a different substrate binding mode specific for the endo- or exonucleolytic reaction. With these equations and the kinetic parameters shown in Figs. 4 and 5, the $K_j$ value (the binding constant) and the $k_{cat}/K_m$ value (identical with the $k_{cat}/K_p$ value) for both the flap and nick substrates were evaluated and compared among mutant and WT enzymes as shown in Fig. 6A, I and II. For the endo- and exo-type hydrolysis, $K_j$ values were reduced remarkably by the replacement at four substrate binding sites (ARG40 and ARG42, Arg118-Lys119, Lys193-Arg194-Lys195) because the $k_{cat}$ was postulated as invariant, strongly indicating that the productive forms partially overlapped for both substrates using all four binding sites. For the endo-type hydrolysis, the mutation at two sites located on the top and bottom sides of the active cleft (Arg94 and Lys193-Arg194-Lys195), induced more remarkable decreases in the $K_p$ value than those located in the middle of the cleft (ARG40 and ARG42, Arg118-Lys119) as shown in Fig. 6A, I and B. Because the active center (orange ball) was observed in the interior of the active cleft of phFEN-1 as shown in Fig. 6B, the hinge of the flap strand might be accessible to the active center after bending at the middle of the substrate. The two positively charged sites (Arg94 and Lys193-Arg194-Lys195) in the top and bottom portions of the phFEN-1 molecule are probably responsible for bending the substrate with their high affinity to DNA.

For the exo-type hydrolysis as shown in Fig. 6A, II, the $K_p$ value was more markedly reduced by mutation in four parts of the active cleft (Arg94, Arg118-Lys119, Arg40 and Arg42, and Lys193-Arg194-Lys195). These facts indicate that all four binding sites play critical roles. The role of Arg118-Lys119 is dominant in the exo-type reaction, in contrast to its less important part in the endo-type reaction. As shown in Fig. 6A, II, the $K_j$ value of the R94A was not significantly changed by the mutation. This clearly indicates that, in the exo-type reaction, Arg94 prefers to take a productive form rather than bind to the substrate, whereas both functions of Arg94 are needed in the endo-type reaction.

The affinity of phFEN-1 for the dsDNA was evaluated by surface plasmon resonance measurement as shown in Table I and Fig. 6A, III. The results indicate that two sites Arg40 and

$$1/K_m = \Sigma K_i$$  
(Eq. 1)

$$k_{cat} = k_{cat}(K_p/\Sigma K_i)$$  
(Eq. 2)

$$k_{cat}/K_m = k_{cat}/K_p$$  
(Eq. 3)
Arg\textsuperscript{42} and Arg\textsuperscript{118}-Lys\textsuperscript{119} mainly contribute to the binding to dsDNA. Comparing the binding affinity of the mutants between Fig. 6A, III and II in which the affinity was evaluated as the \(K\) value, the profiles seems to be similar. The similarity suggested that the binding affinity for both the nicked substrate and dsDNA is produced by two commonly used sites (Arg\textsuperscript{40} and Arg\textsuperscript{42} and Arg\textsuperscript{118}-Lys\textsuperscript{119}) with less of a contribution by the other two sites (Arg\textsuperscript{34} and Lys\textsuperscript{191}, Arg\textsuperscript{194}-Lys\textsuperscript{195}). Interestingly, this indicates that the major binding form might be shared between dsDNA and the nick substrate.

Here, we successfully identified four DNA binding sites that share affinity to form productive ES complexes specific for each endo- or exo-type hydrolysis, probably by bending the substrates. The molecular structure of phFEN-1 was solved as a dimer form; however, unfortunately, we could not obtain the strong biochemical evidence that the dimer formation was necessary to fulfill the function of phFEN-1. A structural analysis of the ES complex with a substrate analogue will clarify the functional role of the dimer formation and the defined mechanism for endo/exonuclease-like sites recognition.

**Acknowledgments**—We thank Y. Shen, H. Higashibata, E. Yokoyama, and S. Tang, for insightful discussions. We are also grateful to E. Yamamoto for providing valuable technical assistance.

**REFERENCES**

1. Gonzalez, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaka, J., and Kato, C. (1998) *Extremophiles*, 2, 123–130
2. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hossyama, A., Nagai, Y., Sakai, M., Oikura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Fuhnahashi, T., Tanaka, T., Kushy, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Yoshizawa, T., Nakamura, Y., Robb, F. T., Horikoshi, K., Matsugi, Y., Shiraya, H., and Kikuchi, H. (1998) *DNA Res.*, 5, 55–76
3. Myllykallio, H., Lopez, P., Garcia, P. L., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000) *Science*, 288, 2212–2215
4. Tye, B. K. (2000) *Proc. Natl. Acad. Sci. U. S. A.*, 97, 2399–2401
5. Kelman, Z. (2000) *Trends. Biochem. Sci.*, 25, 521–523
6. Turchi, J. J., Huang, L., Murante, R. S., Kim, Y., and Bambara, A. (1994) *Proc. Natl. Acad. Sci. U. S. A.*, 91, 8003–8007
7. Murante, R. S., Rumbaugh, J. A., Barnes, C. J., Norton, J. R., and Bambara, R. A. (1996) *J. Biol. Chem.*, 271, 25888–25897
8. Rumbaugh, J. A., Murante, R. S., Shi, S., and Bambara, R. A. (1997) *J. Biol. Chem.*, 272, 22501–22509
9. Bambara, R. A., Murante, R. S., and Henrickson, L. A. (1997) *J. Biol. Chem.*, 272, 4647–4650
10. Khunglaid, A., and Lindahl, T. (1997) *EMBO J.*, 16, 3341–3348
11. Kim, K., Biade, S., and Matsumoto, Y. (1998) *J. Biol. Chem.*, 273, 8842–8848
12. Gary, R., Kim, K., Cornelius, H. L., Park, M. S., and Matsumoto, Y. (1999) *J. Biol. Chem.*, 274, 4354–4363
13. Wu, X., Wilson, T. E., and Lieber, M. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.*, 96, 1303–1308
14. Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A., and Shen, B. (1998) *J. Biol. Chem.*, 273, 27154–27161
15. Rao, H. V. G., Rosenfeld, A., and Wetmur, J. G. (1998) *J. Bacteriol.*, 180, 5406–5412
16. Kaisers, M. W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L., and Lyamichev, Y. I. (1999) *J. Biol. Chem.*, 274, 21317–21319
17. Matsui, E., Kawasaki, S., Ishida, H., Ishikawa, K., Kosugi, Y., Kikuchi, K., Kawarabayashi, Y., and Matsui, I. (1999) *J. Biol. Chem.*, 274, 18297–18309
18. Hwang, K. Y., Baek, K., Kim, H. Y., and Cho, Y. (1998) *Nature Struct. Biol.*, 5, 707–713
19. Hosfield, D. J., Mol, C. D., Shen, B., and Tainer, J. A. (1998) *Cell*, 95, 135–146
20. Kim, Y., Eom, H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995) *Nature*, 376, 612–616
21. Mueser, T. C., Nossal, N. G., and Hyde, C. C. (1996) *Cell J.*, 85, 1101–1112
22. Ceska, T. A., Sayers, J. R., Stier, G., and Suck, D. (1996) *Nature*, 382, 90–93
23. Murante, R. S., Rust, L., and Bambara, R. A. (1995) *J. Biol. Chem.*, 270, 30377–30383
24. Wu, X., Li, J., Li, X., Hsieh, C. L., Burgers, M. J., and Lieber, M. R. (1996) *Nucleic Acids Res.*, 24, 2036–2043
25. Barnes, C. J., Wahl, A. P., Shen, B., Park, M. S., and Bambara, R. A. (1996) *J. Biol. Chem.*, 271, 29624–29631
26. Bornarth, C. J., Ranalli, T. A., Henrickson, L. A., Wahl, A. F., and Bambara, R. A. (1999) *Biochemistry*, 38, 13347–13354
27. Henrickson, L. A., Tom, S., Liu, Y., and Bambara, R. A. (2000) *J. Biol. Chem.*, 275, 16420–16427
28. Harrington, J. J., and Lieber, M. R. (1994) *EMBO J.*, 13, 1235–1246
29. Murante, R. S., Huang, L., Turchi, J. J., and Bambara, R. A. (1994) *J. Biol. Chem.*, 269, 1191–1196
30. Shen, B., Nolan, J. P., Sklar, L. A., and Park, M. S. (1996) *J. Biol. Chem.*, 271, 9173–9176
31. Shen, B., Nolan, J. P., Sklar, L. A., and Park, M. S. (1997) *Nucleic Acids Res.*, 25, 3332–3338
32. Frank, G., Qu, J., Sumsouk, M., Weng, Y., Samsouk, L., Nolan, J. P., and Shen, B. (1998) *J. Biol. Chem.*, 273, 33064–33072
33. Garforth, S. J., Ceska, T. A., Suck, D., and Sayers, J. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.*, 96, 38–43
34. Sakoda, M., and Hirono, K. (1976) *J. Biochem.*, 80, 547–555
35. Kim, C. Y., Park, M. S., and Dyer, R. B. (2001) *Biochemistry*, 40, 3208–3214
36. Hirono, K. (1970) *Biochem. Biophys. Res. Commun.*, 40, 1–6
37. Thoma, J., Brother, C., and Spradlin, J. (1970) *Biochemistry*, 9, 1768–1775
38. Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S., and Honda, K. (1992) *Biochemistry*, 31, 5232–5236
39. Matsui, I., Ishikawa, K., Matsui, E., Miyairi, S., Fukui, S., and Honda, K. (1991) *J. Biochem.* (Tokyo) 100, 566–569
40. Matsui, I., and Svensson, B. (1997) *J. Biol. Chem.*, 272, 22456–22463