A Comparison of Eubacterial and Archaeal Structure-specific 5'-Exonucleases*

(Received for publication, February 26, 1999, and in revised form, April 14, 1999)

Michael W. Kaiser‡, Natasha Lyamicheva, Wupo Ma, Carolyn Miller§, Bruce Neri, Lance Fors, and Victor I. Lyamichev
From Third Wave Technologies, Inc., Madison, Wisconsin 53719

The 5'-exonuclease domains of the DNA polymerase I proteins of Eubacteria and the FEN1 proteins of Eukarya and Archaea are members of a family of structure-specific 5'-exonucleases with similar function but limited sequence similarity. Their physiological role is to remove the displaced 5' strands created by DNA polymerase during displacement synthesis, thereby creating a substrate for DNA ligase. In this paper, we define the substrate requirements for the 5'-exonuclease enzymes from Thermus aquaticus, Thermus thermophilus, Archaeoglobus fulgidus, Pyrococcus furiosus, Methanococcus jannaschii, and Methanobacterium thermoautotrophicum. The optimal substrate of these enzymes resembles DNA undergoing strand displacement synthesis and consists of a bifurcated downstream duplex with a directly abutted upstream duplex that overlaps the downstream duplex by one base pair. That single base of overlap causes the enzymes to leave a nick after cleavage and to cleave several orders of magnitude faster than a substrate that lacks overlap. The downstream duplex needs to be 10 base pairs long or greater for most of the enzymes to cut efficiently. The upstream duplex needs to be only 2 or 3 base pairs long for most enzymes, and there appears to be interaction with the last base of the primer strand. Overall, the enzymes display very similar substrate specificities, despite their limited level of sequence similarity.

* This work was supported by the Cooperative Agreements 70NANB5H1030 and 70NANB7H3015 from the Department of Commerce Advanced Technology Program (to L. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ To whom correspondence should be addressed: Third Wave Technologies, Inc., 502 S. Rosa Rd., Madison, WI 53719. Tel.: 608-273-8933; Fax: 608-273-8618; E-mail: mkaiser@twt.com.
§ Present address: Nanogen, Inc., San Diego, CA 92121.
enzymes is reduced by modifications of the 3′ end of the upstream primer implying recognition of the end of the upstream primer duplex. All enzymes leave a ligatable nick upon cleavage, contrary to published data on identical and related 5′-exonucleases. These data demonstrate that the 5′-structure-specific nucleases are able to carry out the final step of DNA replication prior to ligation.

EXPERIMENTAL PROCEDURES

Materials—PCR7 amplification was done with the Advantage cDNA PCR kit (CLONTECH). Mutagenesis was done with the Transformer site-directed mutagenesis kit (CLONTECH). Restriction enzymes were purchased from New England Biolabs. Chemicals and buffers were from Fisher unless otherwise noted.

Cloning, Expression, and Purification of Enzymes—T. thermophilus strain HB-8 (ATCC 27634) and T. aquaticus strain YT-1 (ATCC 25104) genomic DNAs were used as templates to amplify by PCR the corresponding DNA polymerase I genes, ThpPol and TaqPol, as well as the 5′ nuclelease domain of TaqPol, TaqExo (31). A. fulgidus (DSMZ 4304), M. jannaschii (DSMZ 2661), and M. thermoautotrophicum (ATCC 29096) genomic DNAs were used for PCR amplification of archaeal FEN1 genes, AfuFEN, MjaFEN, and MthFEN, respectively. The T. furiosus FEN1 gene, PfuFEN, was PCR amplified from a genomic clone generously supplied by Dr. Frank Robb (University of Maryland, Baltimore, MD).

The amplified genes were cloned into the expression vector pTrex99a (Amersham Pharmacia Biotech) by standard techniques. Six-amino acid histidine tags were added onto the carboxyl termini of all enzymes by site-directed mutagenesis (TaqPol, TthPol, TaqExo, and PfuFEN) or by including the His tag sequence in the oligonucleotide used for PCR (Mja, Mth, and Afu FEN1 genes). A conserved aspartic acid at position 785 of TaqPol and position 787 of TthPol was mutated to an asparagine by including the His tag sequence in the oligonucleotide used for PCR (Mja, Mth, and Afu FEN1 genes). A conserved aspartic acid at position 785 of TaqPol and position 787 of TthPol was mutated to an asparagine to create polymerase-deficient versions of the enzymes used in this study.

For expression, plasmids were transformed into the E. coli strain BL21 (Novagen), which is deficient in the lon and ompT proteases. Log phase cultures of BL21 were induced with 0.5 mM isopropyl-1-thio-

RESULTS

Cloning and Purification of Seven Structure-specific 5′ Nucleases—Four archaeal FEN1 enzymes from A. fulgidus (AfuFEN), P. furiosus (PfuFEN), M. jannaschii (MjaFEN), and M. thermoautotrophicum (MthFEN), two eubacterial polymerase I enzymes from T. aquaticus (TaqPol) and T. thermophilus (TthPol) and the 5′ nuclease domain of TaqPol (TaqExo) (31) were cloned, expressed in E. coli, and extensively purified to study their substrate specificity (Fig. 1). All enzymes used in this study have six-histidine tags on their carboxyl termini to facilitate purification. To determine whether this modification affects enzyme specificity, five of the enzymes studied here (except TaqExo and MthFEN) were cloned without the histidine tags and were purified by heparin affinity chromatography to a similar level of purity. These enzymes have no measurable differences in specificity compared with their His tag containing counterparts (data not shown). TaqPol and TthPol each contain a single amino acid substitution of aspartic acid to asparagine (D785N for TaqPol and D787N for TthPol) in their polymerase domain, which eliminates polymerization activity; the equivalent mutation in E. coli DNA polymerase I has been shown to have no effect on DNA binding (19) and also has no effect on the cleavage rate or substrate specificity of TaqPol or TthPol (data not shown).

1 The abbreviations used are: PCR, polymerase chain reaction; HPLC, high pressure liquid chromatography; MOPS, 4-morpholinopropanesulfonic acid; bp, base pair(s).
The presence of an upstream primer is known to be important in stimulating cleavage of 5’ arm containing substrates by the structure-specific nucleases (2). Furthermore, overlap between the upstream and downstream duplexes stimulates cleavage of some nucleases even further (17, 20). The identity of the 3’-terminal nucleotide of the upstream primer may also play a role in the stimulation of cleavage (31). To investigate the role of the 3’-terminal nucleotide of the upstream primer in substrate recognition, we created substrates that differed in the identity at that position. In all substrates, the downstream oligonucleotide was labeled with fluorescein at its 5’ end and connected to the template strand via the exceptionally stable GAA hairpin loop (21). Quantitative HPLC analysis of the products of cleavage reactions indicates that the presence of fluorescein at the 5’ end of substrates increases slightly the cleavage rate compared with unlabeled substrates for the enzymes used in this study (data not shown). The 3’ arm of the hairpin is free to anneal to an upstream primer to form the flap structures shown in Fig. 2A.

Reactions with all seven enzymes were performed in the presence of an excess of substrate (2 μM) over enzyme (0.35 nM for TaqExo and FENs and 2.8 nM for TaqPol and TthPol). These substrate concentrations are much higher than the $K_m$ value for all studied substrates (data not shown) assuming that the cleavage rates measured as described under “Experimental Procedures” are close to $V_{max}$ for each enzyme. For all substrate enzyme combinations, the one major product observed corresponded to the product generated by cleavage after the first base pair of the downstream duplex. Cleavage generated five-nucleotide fragments, as shown in Fig. 2B for AfuFEN and TthPol.

For all enzymes a natural base at the 3’ end of the upstream primer supported the highest rate of cleavage (Fig. 2B, lanes 5–8 and 17–20). The archaeal FEN1 enzymes used all four natural bases with approximately equal efficiency, but the cleavage rates of the eubacterial enzymes were clearly dependent on the nature of the 3’-terminal nucleotide. For TaqPol and TaqExo, dT inhibited cleavage, whereas dA supported the highest level of cleavage among the natural bases; for TthPol, dA and dG inhibited cleavage compared with dT and dC. A 3’-terminal dA inhibited cleavage by TaqPol by 30% compared to dT and dG by 60% compared to the nonoverlapping substrate that lacks overlap between the two duplexes, like those previously described in the literature (5).

Reactions with all seven enzymes were performed in the presence of an excess of substrate (2 μM) over enzyme (0.35 nM for TaqExo and FENs and 2.8 nM for TaqPol and TthPol). These substrate concentrations are much higher than the $K_m$ value for all studied substrates (data not shown) assuming that the cleavage rates measured as described under “Experimental Procedures” are close to $V_{max}$ for each enzyme. For all substrate enzyme combinations, the one major product observed corresponded to the product generated by cleavage after the first base pair of the downstream duplex. Cleavage generated five-nucleotide fragments, as shown in Fig. 2B for AfuFEN and TthPol. Release of a five-nucleotide arm should create a nick between the upstream and downstream primers rather than a gap or an overlap. Cleavage rates for all enzymes are summarized in Table I.
phosphate or d-spacer group largely eliminated cleavage for all enzymes. The dideoxy C greatly inhibited the activity of all archaeal enzymes, whereas it reduced the level of cleavage for the eubacterial enzymes by only 10–30% relative to a substrate with dC at the end of the primer strand (Table I). Under these conditions, no cleavage was observed in the absence of an upstream primer (Fig. 2B, lanes 3 and 15).

For all enzymes the substrates with a nonoverlapping flap had a lower rate of cleavage than the substrate with an overlapping natural 3′ nucleotide. Surprisingly, the observed cleavage rate was higher than that reported for cleavage of a similar nonoverlapping type of substrate (17). We hypothesize, as previously suggested (31), that alternative flap structures can be generated to produce substrates with an overlap. In particular, one of the three consecutive As of the template strand could be bulged out to create overlapping structures by slippage (Fig. 2C). To test this hypothesis the sequence of the template strand was modified to prevent the potential slippage by substituting a G-C for an A-T in the middle of the AAA track as shown in Fig. 3A, and the overlapping and nonoverlapping flap substrates were incubated with each of the seven enzymes in reaction conditions identical to those in Fig. 2B. As shown in Fig. 3B, cleavage of the nonoverlapping flap substrate was almost undetectable under the background level (no enzyme control) for all enzymes except AfuFEN, which had only 5% of the activity observed for the overlapping substrate (Fig. 3B, lanes 9 and 13). We conclude that overlap is required for efficient cleavage and that slippage can occur to create overlaps. It is unclear whether this slippage is induced by the enzyme itself.

**Overlapping and Hairpin Substrates and the Effect of Mg and Mn Ions on Enzyme Activity**—After establishing the key features of substrate recognition we designed two substrates to further investigate the specificity of the seven nucleases. The first (overlapping flap substrate, Fig. 4A) has upstream and downstream strands connected to the template strand by two GAA loops to form a “dumbbell” structure. The upstream and downstream duplexes overlap by 3 base pairs. The second substrate (hairpin substrate, Fig. 4B) was designed to study cleavage in the absence of the upstream primer; it has the same 5′ arm and substrate duplex region connected to the short 3′ TTT arm. To obtain similar levels of cleavage, the incubation time and enzyme concentration were varied as shown in Fig. 4.

The cleavage patterns for all enzymes are shown in Fig. 4, and the corresponding cleavage rates are summarized in Table II for both the overlapping and hairpin substrates in MgCl₂ and MnCl₂. The most striking aspect of these data is the increase in cleavage rate and cleavage accuracy conferred by the presence of the upstream primer duplex. For instance, in MgCl₂, the presence of the upstream primer duplex caused the cleavage rate of TaqExo to increase 4 orders of magnitude. An even greater increase is seen for AfuFEN and PfuFEN, because cleavage of the hairpin substrate cannot be detected in MgCl₂.

**TABLE I**

| Substrate with no overlap (Fig. 2A) | dA | dT | dG | dC | ddC | dCp | d-spacer |
|-----------------------------------|----|----|----|----|-----|-----|----------|
| TaqExo 3.7 ± 0.8                  | 61.5 ± 1.0 | 16.1 ± 0.3 | 43.6 ± 1.1 | 49.3 ± 0.3 | 34.1 ± 0.7 | 0.9 ± 0.1 | 1.8 ± 0.1 |
| TaqPol 1.9 ± 0.1                  | 32.8 ± 0.6 | 9.2 ± 0.5 | 27.3 ± 0.2 | 25.4 ± 0.2 | 19.1 ± 0.5 | 1.5 ± 0.2 | 0.9 ± 0.1 |
| ThhPol 7.8 ± 1.9                  | 47.7 ± 0.7 | 77.7 ± 2.8 | 48.3 ± 1.4 | 82.5 ± 3.4 | 72.7 ± 4.8 | 2.3 ± 0.3 | 10.4 ± 0.4 |
| AfuFEN 18.4 ± 0.1                 | 84.8 ± 2.5 | 93.9 ± 0.8 | 89.0 ± 0.6 | 92.1 ± 3.1 | 4.0 ± 0.1 | 7.8 ± 0.1 | 8.2 ± 0.1 |
| PfuFEN 36.9 ± 1.9                 | 99.1 ± 6.9 | 122.4 ± 3.6 | 96.0 ± 1.7 | 147.6 ± 3.1 | 4.8 ± 0.3 | 7.0 ± 0.1 | 5.7 ± 0.4 |
| MjaFEN 120.7 ± 18.3              | 277.0 ± 17.5 | 326.9 ± 18.0 | 318.6 ± 8.7 | 290.7 ± 9.4 | 77.4 ± 1.3 | 18.2 ± 0.8 | 90.9 ± 3.7 |
| MthFEN 58.2 ± 9.3                | 224.1 ± 8.1 | 181.2 ± 7.8 | 185.2 ± 9.6 | 189.4 ± 9.1 | 13.9 ± 2.1 | 4.5 ± 0.8 | 15.7 ± 1.2 |

*Substrate with no overlap (Fig. 2A).*

Using the background signal as the upper limit of hairpin cleavage, we estimate that the difference in cleavage rate of two substrates for these enzymes is at least 60,000-fold. Substitution of MnCl₂ for MgCl₂ stimulated cleavage of the hairpin by about 1 or 2 orders of magnitude, depending on the enzyme, whereas it decreased the rate of cleavage of the overlapping flap by about 0.5 order of magnitude. Although the stimulatory effect of the upstream duplex is not as great in the presence of MnCl₂, it is still significant, averaging 1.5–2 orders of magnitude. The differences between Tables I and II in cleavage rates for the overlapping substrates can be explained by differences in the MnCl₂ and KCl concentrations as well as by subtle differences in the structures of the substrates (compare Figs. 2A and 4A).

The products of cleavage of the hairpin substrate differed between enzymes, although the major product of most enzymes is 5 nucleotides long, indicating that cleavage occurs after the first base pair of the substrate duplex. The exception is MthFEN, whose major product was 6 nucleotides long. In addition to significantly increasing the cleavage rate, the presence of the upstream primer duplex also made the cleavage pattern nearly identical for all enzymes. For both the overlapping flap and hairpin substrates, the positions of cleavage were not affected by the choice of divalent cation. The major cleavage product of the overlapping flap substrate is produced by cleavage at the
position expected if the upstream primer is fully base paired. The appearance of a small amount of 3'-terminal nucleotides of the upstream primer unpaired.

**TABLE II**

| Bacterial and Archaeal Structure-specific 5'-Exonucleases | Exonuclease | Overlap Flap Substrate | Hairpin Substrate |
|----------------------------------------------------------|-------------|------------------------|-------------------|
| Activity of 5'-exonucleases on overlapping flap and hairpin substrates in MgCl₂ and MnCl₂ | TaqExo | 194 | 70 | 0.08 | 10 |
| | TaqPol | 3 | 14 | 0.003 | 0.16 |
| | ThhPol | 1 | 6 | a | 0.02 |
| | MjaFEN | 158 | 51 | a | 0.04 |
| | PfuFEN | 195 | 50 | a | 0.03 |
| | MthFEN | 520 | 138 | 0.76 | 68 |
| | MthFEN | 230 | 50 | 0.15 | 2.1 |

* pH values are given at reaction temperature 50 °C.

The cleavage reaction was highly dependent on the pH of the buffer, varying as much as 2 orders of magnitude in the range from pH 6 to 10. Most enzymes had optimal activity between pH 8 and 9, with the exceptions of MthFEN and MjaFEN, which reach maximum activity at pH 10 (Table III).

With the overlapping flap substrate in MgCl₂, archaeal enzymes had fairly sharp temperature optima between 50 and 55 °C, TaqExo and TaqPol enzymes had optimal activity between 65 and 70 °C, respectively, and ThhPol was most active at 85 °C (Table IV). All enzymes were able to cut the overlapping flap substrate at both 40 and 90 °C and survived incubation at 90 °C for 15 min, except MthFEN, which is not thermostable above 75 °C (Table IV and data not shown). All enzymes had lower temperature optima in MnCl₂ than MgCl₂ except for TaqPol and TaqExo, for which the optima were increased by 5–10 °C (data not shown).

**The Effect of Substrate Structure on Cleavage Efficiencies**—Having established a standard set of reaction conditions, we undertook a comparison of how structural elements of the hairpin and overlapping flap substrates affected cleavage by the seven nuclease enzymes. The effect of the 3' arm length in the hairpin substrate was determined with a series of hairpin substrates like the substrate in Fig. 4B but having 3'-arms of 0, 3, 8, or 15T nucleotides. Surprisingly, the 3' arm length can have as great an effect on the cleavage rate. The archaeal FEN1 enzymes cleaved at the highest rate when the 3' arm was 8 nucleotides long, whereas the eubacterial enzymes were most active on substrates lacking a 3' arm (Table V). To determine the minimal length of upstream duplex re-
Activities were measured with the overlap (Fig. 4A) substrate in the presence of 4 mM MgCl₂ as described under “Experimental Procedures” and presented as percentages relative to highest activity.

| Temperature | 40 °C | 50 °C | 55 °C | 60 °C | 65 °C | 70 °C | 75 °C | 80 °C | 85 °C | 90 °C |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TaqExo      | 55    | 82    |       | 97    | 100   | 81    | 60    |       |       |       |
| TaqPol      | 12    | 42    |       | 63    | 97    | 100   | 60    |       |       |       |
| ThhPol      | 1.3   | 4     |       | 13    | 25    | 36    | 60    | 82    | 100   |       |
| AfuFEN      |       |       |       |       | 61    | 77    | 100   | 84    | 33    | 11    |
| PhuFEN      | 11    | 23    |       | 27    | 32    | 80    | 100   | 85    | 28    |        |
| MjaFEN      |       |       |       |       | 33    | 53    | 10    | 85    | 84    | 32    |
| MthFEN      | 10⁷b  | 26/20⁵ | 39/31⁵ | 52/53⁵ | 71/77⁷ | 95/88⁸ | 100/100⁹ | 67/3³ |       |       |

Activities not determined.

TABLE V
Effect of temperature on activity of seven structure-specific 5′-exonucleases

| Temperature | 40 °C | 50 °C | 55 °C | 60 °C | 65 °C | 70 °C | 75 °C | 80 °C | 85 °C | 90 °C |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TaqExo      | 55    | 82    |       | 97    | 100   | 81    | 60    |       |       |       |
| TaqPol      | 12    | 42    |       | 63    | 97    | 100   | 60    |       |       |       |
| ThhPol      | 1.3   | 4     |       | 13    | 25    | 36    | 60    | 82    | 100   |       |
| AfuFEN      |       |       |       |       | 61    | 77    | 100   | 84    | 33    | 11    |
| PhuFEN      | 11    | 23    |       | 27    | 32    | 80    | 100   | 85    | 28    |        |
| MjaFEN      |       |       |       |       | 33    | 53    | 10    | 85    | 84    | 32    |
| MthFEN      | 10⁷b  | 26/20⁵ | 39/31⁵ | 52/53⁵ | 71/77⁷ | 95/88⁸ | 100/100⁹ | 67/3³ |       |       |

TABLE IV
Effect of temperature on activity of seven structure-specific 5′-exonucleases

| Temperature | 40 °C | 50 °C | 55 °C | 60 °C | 65 °C | 70 °C | 75 °C | 80 °C | 85 °C | 90 °C |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TaqExo      | 55    | 82    |       | 97    | 100   | 81    | 60    |       |       |       |
| TaqPol      | 12    | 42    |       | 63    | 97    | 100   | 60    |       |       |       |
| ThhPol      | 1.3   | 4     |       | 13    | 25    | 36    | 60    | 82    | 100   |       |
| AfuFEN      |       |       |       |       | 61    | 77    | 100   | 84    | 33    | 11    |
| PhuFEN      | 11    | 23    |       | 27    | 32    | 80    | 100   | 85    | 28    |        |
| MjaFEN      |       |       |       |       | 33    | 53    | 10    | 85    | 84    | 32    |
| MthFEN      | 10⁷b  | 26/20⁵ | 39/31⁵ | 52/53⁵ | 71/77⁷ | 95/88⁸ | 100/100⁹ | 67/3³ |       |       |

Fig. 5. Effect of upstream duplex length. A, sequence and proposed structures of overlapping flap substrates with 5- and 0-bp upstream duplexes. Upstream duplexes shorter than 5 bp were generated by removing base pairs from the loop side of the duplex. B, substrates with 0-bp (lanes 3, 9, and 15), 1-bp (lanes 4, 10, and 16), 2-bp (lanes 5, 11, and 17) 3-bp (lanes 6, 12, and 18), 4-bp (lanes 7, 13, and 19), and 5-bp (lanes 8, 14, and 20). The upstream duplex was removed with 0.35 mM AfuFEN for 8 min (lanes 3–8), 0.35 mM MjaFEN (lanes 9–14) for 4 min, or 2.8 nM TthPol (lanes 15–20) for 4 min with 4 mM MgCl₂, as described under “Experimental Procedures.” Lanes 1 and 2, size markers with sequences identical to the 5′-terminal sequence of the primer; lane 2, no enzyme control.

The effect of downstream duplex length on cleavage efficiency was determined using substrates with downstream duplex lengths of between 8 and 16 base pairs for the hairpin substrate and 6 and 16 base pairs for the overlapping flap substrate (Fig. 6A), in which the upstream duplex was 6 base pairs. All enzymes, except TaqPol, cleaved the overlapping flap substrates at rates independent of downstream duplex length in the range from 10 to 16 base pairs; TaqPol cleaved the 10 base pair substrate approximately five times slower than the 12 base pair substrate (Fig. 6B). For most enzymes, cleavage activities decreased for the 8 base pair downstream duplex substrate and significantly dropped when the duplex length was reduced to 6 bp. Only MjaFEN and MthFEN were able to cleave the substrate with the 6-base pair downstream duplex. The hairpin substrates, cleavage rates decreased nearly linearly with decreasing duplex length and were only 5–10% of maximal (data not shown) with an 8-base pair hairpin for TthPol and the FEN enzymes. TaqPol and TaqExo had essentially no activity on that substrate.

Ligation after Cleavage by the Structure-specific 5′-Exonucleases—In vivo, strand displacement to remove RNA primers or damaged DNA should ultimately generate a structure that can be sealed by ligation. To determine whether the enzymes studied here are capable of generating such nicked duplexes, we determined whether the upstream and downstream primers could be ligated after cleavage in the absence of any polymerase activity. Ligation would also confirm that the deduced site of cleavage is juxtaposed to the 3′ end of the upstream primer. In these experiments, the substrate was generated by annealing 3 oligonucleotides and cleaving the resulting substrate with each of the enzymes followed by the addition of T4 DNA ligase. The oligonucleotide that was to be cleaved was labeled at its 3′ end to permit monitoring of its ligation to the upstream primer oligonucleotide (Fig. 7). The 3′ labeled substrate was cleaved at approximately the same rate as the 5′ labeled substrate for all seven enzymes within experimental error (data not shown),
substrates described in pairs from the loop side of the duplex. Plexes were generated by removing base with 16-bp downstream duplex. Substructure of overlapping flap substrate substrate. length on cleavage of overlapping 0.35 nM TaqExo for 8 min, 2.8 nM TaqPol for 3 min, 2.8 nM TthPol for 2 min, 0.35 nM AfuFEN for 8 min, 0.35 nM PfuFEN for 6 min, 0.35 nM MjaFEN for 4 min, or 0.35 nM MthFEN for 4 min with 4 mM MgCl₂ as described under “Experimental Procedures.” The cleavage rates are plotted versus the downstream duplex length, each point represents the average of three data points, and errors are shown by vertical lines.

DISCUSSION

During DNA replication, a large number of RNA primers must be removed, and the Okazaki fragments generated by those RNA primers must be joined by DNA ligase. The literature regarding the question of which enzyme acts immediately prior to DNA ligase is contradictory. The work by Lundquist and Oliviera (1) demonstrated the ability of E. coli DNA polymerase I to leave a nick between the upstream and downstream strands after cleavage of displaced single-stranded overhangs generated during nick translation. However, studies with the whole DNA polymerase I or its isolated 5'-exonuclease domain only gave a different result (13). Using a preformed substrate, the nuclease leaves a gap, leading the authors to speculate that the DNA polymerase must then act to fill in that gap to generate a ligatable nick. A number of other 5'-nucleases have been shown to leave a gap or overlap after cleavage of the same or similar flap substrates (5, 8, 22, 23).

The principal finding of this work is that all the structure-specific 5'-exonucleases leave a nick after cleavage of a substrate that has the overlap between the upstream and downstream duplexes (Figs. 2 and 3). When overlap exists, there can be branch migration of the two duplexes resulting in the 3' end of the upstream primer being alternately paired and unpaired (1, 17). Our data show that cleavage occurs on the substrate having the conformation where the last nucleotide at the 3' end of the upstream strand is unpaired because the cleavage rate is essentially the same whether the end of the upstream primer is A, C, G, or T (Fig. 2). Thus, it is positional overlap between the 3' end of the upstream primer and downstream duplex rather then sequence overlap that is required for optimal cleavage. The fact that the 3' end of the upstream primer is unpaired indicates that, as in the case of the hairpin substrates (2), the cleavage occurs between the first two base pairs of the downstream duplex. Its also explains why nonoverlapping flap substrates, such as described (5), generate gaps. When the upstream duplex abuts the downstream duplex with no overlap, cleavage after the first base pair of the downstream duplex (described as proximal in Ref. 5) will create a gap.

Comparing the results from Figs. 2 and 3, it appears that the enzymes are able to tolerate slippage of the two strands of the upstream duplex to generate overlap and thus a more cleavable substrate (Fig. 2C), even when such slippage would cause a distortion in the DNA duplex. Alternative structures of the downstream duplex that create overlap with the upstream primer have been proposed to explain cleavage of nonoverlapping substrates by TaqPol and TaqExo. We believe the ability of the 5'-nucleases to support cleavage of the nonoverlapping flap structure used by Harrington and Lieber (5) may be due to a similar rearrangements of the downstream duplex (distal cutting in Ref. 5). The fact that mutations that stabilize the 3' end of the downstream duplex significantly reduce cleavage for some enzymes supports that idea (8).

Another aspect of cleavage affected by overlap is the cleavage rate. We can directly compare our results for PfuFEN, AfuFEN, and MjaFEN with data from the literature because cleavage under “Experimental Procedures.” After cleavage, the reactions were incubated at 23 °C with or without T4 DNA ligase for 15 min, and the products were resolved on a 15% polyacrylamide gel. The following 5'-nucleases were used; lanes 1 and 2, no 5'-nuclease; lanes 3 and 4, AfuFEN; lanes 5 and 6, PfuFEN; lanes 7 and 8, MjaFEN; lanes 9 and 10, MthFEN; lanes 11 and 12, TaqExo; lanes 13 and 14, TagPol; lanes 15 and 16, TthPol. Even-numbered lanes contained 1 unit of T4 DNA ligase. Lane 17 contains an oligonucleotide with the expected sequence of the ligation product.

Fig. 6. Effect of downstream duplex length on cleavage of overlapping substrate. A, sequence and proposed structure of overlapping flap substrate with 16-bp downstream duplex. Substrates with shorter downstream duplexes were generated by removing base pairs from the loop side of the duplex. B, substrates described in A were cut with 0.35 nM TaqExo for 8 min, 2.8 nM TaqPol for 3 min, 2.8 nM TthPol for 2 min, 0.35 nM AfuFEN for 8 min, 0.35 nM PfuFEN for 6 min, 0.35 nM MjaFEN for 4 min, or 0.35 nM MthFEN for 4 min with 4 mM MgCl₂ as described under “Experimental Procedures.” The cleavage rates are plotted versus the downstream duplex length, each point represents the average of three data points, and errors are shown by vertical lines.

Fig. 7. Ligation of cleaved products. Cleavage reactions were carried out for 5 min at 50 °C on a substrate comprised of three separate oligonucleotides that form an overlapping flap substrate as described under “Experimental Procedures.” After cleavage, the reactions were incubated at 23 °C with or without T4 DNA ligase for 15 min, and the products were resolved on a 15% polyacrylamide gel. The following 5'-nucleases were used; lanes 1 and 2, no 5'-nuclease; lanes 3 and 4, AfuFEN; lanes 5 and 6, PfuFEN; lanes 7 and 8, MjaFEN; lanes 9 and 10, MthFEN; lanes 11 and 12, TaqExo; lanes 13 and 14, TagPol; lanes 15 and 16, TthPol. Even-numbered lanes contained 1 unit of T4 DNA ligase. Lane 17 contains an oligonucleotide with the expected sequence of the ligation product.
0.04 nM/s combined for all products (19 and 21-mer) were observed for MjaFEN, AfuFEN, and PfuFEN, respectively. When converted to the cleavage rates as described under “Experimental Procedures,” those numbers are equivalent to 0.032, 0.015, and 0.003 cleavages/minute for MjaFEN, AfuFEN, and PfuFEN, respectively. These rates are more than 3 orders of magnitude slower than the rates we observe for the same enzymes with the overlapping flap substrate. Part of the difference may be explained by differences in methodology and in the particular substrate used. However, a cleavage rate decrease of several orders of magnitude is consistent with what we observe for most enzymes except AfuFEN. The cleavage rate of AfuFEN only decreases a single order of magnitude when overlap is eliminated for the substrate used in Fig. 3. Further experimentation will be required to determine whether this result is reproducible with different substrates. The stimulatory effect of overlap has been reported for human FEN1 (20), PfuFEN and AfuFEN (17), and TaqPol (31), but quantitative analysis of this effect has not been done. Interestingly, length of the overlap between the two duplexes seems to not effect the cleavage rate (Refs. 17 and 20 and data not shown), which indicates that branch migration is fast compared with enzyme binding and cleavage.

Although the upstream primer stimulates and focuses cleavage by the 5′ nucleases, the downstream duplex alone is sufficient to support cleavage by the 5′ nucleases (Fig. 4B). The three-dimensional structures for five members of the 5′ nuclease family have been solved (24–28), and all five show two divalent metal binding sites. It has been proposed for human FEN1 (29) and for T4 RNase H (30) that one metal ion is involved primarily in binding to the DNA substrate. It has also been proposed that the structure-specific 5′-nucleases contain the helix-hairpin-helix motif, which is thought to bind to double-stranded DNA in a nonspecific manner (16). When we examined the three-dimensional structure of Taq polymerase determined by x-ray diffraction (24), we noticed that the distance between the metal ion binding site and the helix-hairpin-helix motif (aspartic acid 142 to lysine 197) is about 30 Å or roughly 9 base pairs of B-form helix. A contact length of 12 base pairs between PfuFEN and a DNA substrate was recently proposed based on the length of a prominent groove on the surface of that enzyme (28). It is interesting to note that cleavage by TaqExo drops 10-fold going from a 10- to an 8-base pair downstream duplex length (Fig. 6). We hypothesize that one of the bound divalent cations and the helix-hairpin-helix motif are the primary binding sites for the downstream duplex of DNA substrates on the 5′ nucleases.

The presence of the upstream duplex is important for both activating and positioning cleavage by the structure-specific 5′-exonucleases (Fig. 2), but only a small portion of that duplex is recognized by the enzymes (Fig. 5). Most enzymes still show a significant amount of cleavage when the upstream duplex is as small as two base pairs, although the stability of this duplex at 50 °C is questionable. In particular, recognition of the last nucleotide of the upstream primer strand appears to be important. Removal of a single oxygen atom by substitution of a deoxy-ribose with a deoxy-ribose at the 3′ end of the primer duplex significantly reduces cleavage by archaeal enzymes, and substitution of a 3′ OH with a 3′ phosphate affects all enzymes (Fig. 2 and Table I). Coupled with the fact that the 3′ nucleotide of the upstream primer is unpaired during cleavage, this leads us to propose that there is a region or pocket in the enzymes that specifically recognizes that nucleotide. Co-crystallization or other structure determination studies will be needed to identify this region and to elucidate the molecular basis of enzyme recognition.

Many of the aspects of substrate recognition are identical or very similar between the enzymes examined here. We have observed the cleavage rate increase caused by overlap between the upstream and downstream duplexes of the flap substrate for all enzymes examined and using many different substrate sequences (Ref. 17 and data not shown). Furthermore, the cleavage rate does not depend on the GC or AT content of the substrates and is remarkably constant between different substrates (data not shown). We have also found that all enzymes leave a nick after cleavage when overlap exists. We expect these conclusions to be true for all other 5′ structure-specific exonucleases because the members of the group examined here are evolutionarily distant from one another, spanning two kingdoms. Further studies will be needed to determine whether in fact that is true.

Acknowledgments—We thank Frank Robb for providing us with a clone containing the Pfu FEN1 gene. We also thank Jim Dahlberg and Mary Ann Brow for critically reading the manuscript.

REFERENCES
1. Lundquist, R. C., and Oliveira, B. M. (1992) Cell 31, 53–60
2. Lyamichev, V., Brow, M. A., and Dahlberg, J. E. (1993) Science 260, 778–783
3. Kornberg, A., and Baker, A. (1992) DNA Replication, 2nd Ed., pp. 403–437, Freeman, New York
4. Lindahl, T., Gally, J. A., and Edelman, G. M. (1969) Proc. Natl. Acad. Sci. U. S. A. 62, 597–603
5. Harrington, J. J., and Lieber, M. R. (1994) EMBO J. 13, 1235–1246
6. Murante, R. S., Huang, L., Turchi, J. J., and Bambara, R. A. (1994) J. Biol. Chem. 269, 1191–1196
7. Robins, P., Pappin, D. J., Wood, R. D., and Lindahl, T. (1994) J. Biol. Chem. 269, 28535–28538
8. Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A., and Shen, B. (1998) J. Biol. Chem. 273, 27154–27161
9. Rao, H. G., Rosenfeld, A., and Wetmur, J. G. (1998) J. Bacteriol. 180, 5406–5412
10. Bhagwat, M., Hobbs, L. J., and Nossal, N. G. (1997) J. Biol. Chem. 272, 28523–28530
11. Garforth, S. J., and Sayers, J. R. (1997) Nucleic Acids Res. 25, 3801–3807
12. Harrington, J. J., and Lieber, M. R. (1994) Genes Dev. 8, 1344–1355
13. Xu, Y., Derbyshire, V., Ng, K., Sun, X. C., Grindley, N. D., and Joyce, C. M. (1997) J. Mol. Biol. 268, 284–302
14. Bambara, R. A., Murante, R. S., and Henriksson, L. A. (1997) J. Biol. Chem. 272, 4457–4460
15. Turchi, J. J., Huang, L., Murante, R. S., Kim, Y., and Bambara, R. A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 9803–9807
16. Doberty, A. J., Serpell, L. C., and Ponting, C. P. (1996) Nucleic Acids Res. 24, 2488–2497
17. Lyamichev, V., Mael, A. L., Hall, J. G., Prudent, J. R., Kainer, M. W., Takova, T., Kwiatkowski, R., Sanders, T., de Arruda, M., Arco, D., Neri, B. P., and Bow, M. A. D. (1996) Nature BioTechnology 17, 292–296
18. Gill, S. C., and von Hippel, P. H. (1989) Anal. Biochem. 182, 319–326
19. Polesky, A. H., Steitz, T. A., Grindley, N. D., and Joyce, C. M. (1990) J. Biol. Chem. 265, 14579–14591
20. DeMott, M. S., Shen, B., Park, M. S., Bambara, R. A., and Zigmam, S. (1996) J. Biol. Chem. 271, 30068–30076
21. Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K., and Miura, K. (1992) Nucleic Acids Res. 20, 3891–3896
22. Murante, R. S., Rust, L., and Bambara, R. A. (1995) J. Biol. Chem. 270, 30377–30383
23. Pickering, T. J., Garforth, S. J., Thorpe, S. J., Sayers, J. R., and Grashy, J. A. (1999) Nucleic Acids Res. 27, 730–735
24. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995) Nature 376, 612–616
25. Cresca, T. A., Sayers, J. R., Stier, G., and Suck, D. (1996) Nature 382, 90–93
26. Mueser, T. C., Nossal, N. G., and Hyde, C. C. (1996) Cell 85, 1101–1112
27. Hawg, K. Y., Baeck, K., Kim, H. Y., and Cho, Y. (1998) Nat. Struct. Biol. 5, 707–713
28. Hosfield, D. J., Mol, C. D., Shen, B., and Tainer, J. A. (1998) Cell 95, 135–146
29. Shen, B., Nolan, J. P., Sklar, L. A., and Park, M. S. (1997) Nucleic Acids Res. 25, 3332–3338
30. Bhagwat, M., Meara, D., and Nossal, N. G. (1997) J. Biol. Chem. 272, 28531–28538
31. Lyamichev, V., Brow, M. A., Varvel, V. E., and Dahlberg, J. E. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 6143–6148