doi:10.1016/S0022-2836(03)00351-6 J. Mol. Biol. (2003) 328, 537–554

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Modeling of Flap Endonuclease Interactions with DNA Substrate

Hatim T. Allawi[1]*, Michael W. Kaiser[1], Alexey V. Onufriev[2], Wu-Po Ma[1] Andrew E. Brogaard[1], David A. Case[2], Bruce P. Neri[1] and Victor I. Lyamichev[1]

[1]*Third Wave Technologies, Inc. 502 S Rosa Road, Madison, WI 53719, USA*

[2]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037 USA*

**Corresponding author*

Structure-specific 5′ nucleases play an important role in DNA replication and repair uniquely recognizing an overlap flap DNA substrate and processing it into a DNA nick. However, in the absence of a high-resolution structure of the enzyme/DNA complex, the mechanism underlying this recognition and substrate specificity, which is key to the enzyme's function, remains unclear. Here, we propose a three-dimensional model of the structure-specific 5′ flap endonuclease from *Pyrococcus furiosus* in its complex with DNA. The model is based on the known X-ray structure of the enzyme and a variety of biochemical and molecular dynamics (MD) data utilized in the form of distance restraints between the enzyme and the DNA. Contacts between the 5′ flap endonuclease and the sugar–phosphate backbone of the overlap flap substrate were identified using enzyme activity assays on substrates with methylphosphonate or 2′-*O*-methyl substitutions. The enzyme footprint extends two to four base-pairs upstream and eight to nine base-pairs downstream of the cleavage site, thus covering 10–13 base-pairs of duplex DNA. The footprint data are consistent with a model in which the substrate is bound in the DNA-binding groove such that the downstream duplex interacts with the helix-hairpin-helix motif of the enzyme. MD simulations to identify the substrate orientation in this model are consistent with the results of the enzyme activity assays on the methylphosphonate and 2′-*O*-methyl-modified substrates. To further refine the model, 5′ flap endonuclease variants with alanine point substitutions at amino acid residues expected to contact phosphates in the substrate and one deletion mutant were tested in enzyme activity assays on the methylphosphonate-modified substrates. Changes in the enzyme footprint observed for two point mutants, R64A and R94A, and for the deletion mutant in the enzyme's β_A/β_B region, were interpreted as being the result of specific interactions in the enzyme/DNA complex and were used as distance restraints in MD simulations. The final structure suggests that the substrate's 5′ flap interacts with the enzyme's helical arch and that the helix-hairpin-helix motif interacts with the template strand in the downstream duplex eight base-pairs from the cleavage site. This model suggests specific interactions between the 3′ end of the upstream oligonucleotide and the enzyme. The proposed structure presents the first detailed description of substrate recognition by structure-specific 5′ nucleases.

© 2003 Elsevier Science Ltd. All rights reserved

Keywords: structure-specific 5′ nuclease; FEN1; point mutation; methylphosphonate; 2′-*O*-methyl

Introduction

Structure-specific 5′ nucleases have been isolated from different organisms including bacteriophages,[1,2] eubacteria,[3–5] archaea,[6–8] yeast[9,10] and

Abbreviations used: HhH, helix-hairpin-helix.

E-mail address of the corresponding author: hallawi@twt.com

Figure 1. Ribbon diagram of the crystal structure of PfuFEN1.[23] The helix-hairpin-helix motif, the helical arch, the magnesium ions, and the β_A/β_B loop referred to in this work are shown in red, yellow, pink, and orange, respectively. Arginine 64 and 94 are labeled and shown in green.

mammals.[11–14] The ubiquitous presence of these enzymes is explained by their essential role in Okazaki fragment processing[15–17] and repair of DNA damage caused by alkylating agents or UV radiation.[9,18,19] In Okazaki fragment processing, displacement synthesis by DNA polymerases generates branched DNA structures in which the upstream and downstream strands overlap and compete for the same sequence of the template strand. The branched structure can exist in multiple conformations depending on the position of the branch point between the upstream and downstream strands on the shared template sequence.[15,20] The structure-specific 5′ nucleases known as flap endonucleases (FEN1)[12] specifically recognize a conformation called the overlap flap or 3′ one-nucleotide double-flap structure in which the upstream strand, excluding the 3′ end nucleotide, is annealed to the template strand, displacing the 5′ portion of the downstream strand.[8,21,22] FEN1 cleaves the downstream strand of the overlap flap structure precisely after the first base-paired nucleotide, creating a ligatable nick.[8,22]

The conclusion that the 3′ end nucleotide of the upstream strand is not base-paired with the template has been put forth from the observation that any of the four natural bases at this position can support efficient cleavage.[21] It was suggested that the 3′ end nucleotide of the upstream strand interacts with the enzyme to position the substrate in an optimal orientation for cleavage. The demonstration that sugar modifications of the 3′ end nucleotide inhibit activity of FEN1 enzymes has further supported this hypothesis.[8,22] The importance of the overlapping 3′ end nucleotide of the upstream strand was not originally recognized, and many laboratories characterized FEN1 enzymes using a flap structure that included adjacent upstream and downstream strands annealed to a template, but lacking a gap or overlap.[10] FEN1 enzymes cleave such a flap substrate inefficiently, producing a majority of products that are not ligatable. The existence of these products can be explained by the formation of alternative structures with bulged nucleotides stabilized by the 5′ nuclease in an effort to force the overlap flap structure.[8,21,22]

X-ray crystal structures have been determined for six structure-specific 5′ nucleases: the archaeal FEN1 enzymes from *Pyrococcus furiosus* (PfuFEN1)[23] (Figure 1), *Methanococcus jannaschii*

(MjaFEN1)[7] and *Pyrococcus horikoshii* (PhoFEN1);[24] the 5′ nuclease domain of eubacterial DNA polymerase from *Thermus aquaticus* (TaqExo);[25] the 5′–3′ exonuclease from bacteriophage T5,[26] and the RNase H enzyme from bacteriophage T4.[27] These structures reveal a common α/β topology and similar structural motifs despite a low amino acid sequence identity and similarity between the archaeal, eubacterial and bacteriophage FEN1 groups (see, for example, Hosfield *et al.*[23]). All of these enzymes have been shown to bind two divalent metal ions which form a complex network of interactions with highly conserved acidic amino acids lying at the bottom of a positively charged cleft. One metal ion is presumably involved in catalysis and the other in the DNA binding.[28] In the PfuFEN1 structure, the magnesium ion M-1, involved in catalysis, is located in close proximity to the cluster of amino acid residues Asp27, Asp80, Glu152, and Glu154; and the magnesium ion M-2 involved in substrate binding interacts with the cluster of amino acid residues Asp173, Asp175, and Asp236 approximately 5 Å from M-1.

The helical arch is a common structural motif shared by the FEN1 enzymes and was originally identified in the T5 5′–3′ exonuclease structure.[26] The arch is located close to the enzyme's active site and forms a flexible loop that can accommodate single-stranded but not double-stranded DNA. The motif provides structural support for the hypothesis that the 5′ flap of DNA substrate threads through a hole to translocate DNA to the enzyme's active site. The threading mechanism was originally proposed to explain biochemical data that blocking the free 5′ end of the flap with a bulky modification or rendering it double-stranded using a complementary oligonucleotide suppresses the cleavage efficiency and can even trap FEN1 on the 5′ flap.[5,29] While most studies agree on the threading mechanism, Bambara and his group have shown that a variety of bulky flap modifications can be tolerated by human FEN1 endonuclease.[30]

Another common fold shared by the FEN1 enzymes is the helix-hairpin-helix (HhH) motif found in many enzyme families.[31,32] This type of fold is involved in non-sequence-specific binding of duplex DNA *via* interactions with the sugar–phosphate backbone of one of the strands.[33,34] Together with the helical arch and network of amino acids interacting with the M-1 and M-2 ions, the HhH motif defines a positively charged active-site DNA-binding groove in FEN1. In PfuFEN1, the DNA-binding groove is 32 Å wide and 44 Å long, suggesting that it can accommodate a 12 base-pair double-stranded DNA.[23] Biochemical analysis of point mutations at the DNA-binding groove of the FEN1 enzymes revealed conserved amino acid residues on the surface of the groove involved in catalysis and substrate binding.[24,28,35–39]

Structural and functional similarity between the 5′ nucleases suggests a common mechanism for substrate binding and catalysis for all enzymes in this family. In the absence of co-crystal or NMR structures for a 5′ nuclease/DNA complex, several models of the complex have been proposed to elucidate the mechanism of substrate binding.[7,23,26,40] These models suggest that the substrate binds at the active-site DNA-binding groove with the cleavable phosphodiester linkage close to the metal ion involved in catalysis and with the 5′ flap threading through the helical arch.

Methylphosphonate and 2′-*O*-methyl substitutions have proven to be powerful methods for identifying contacts between nucleic acids and proteins.[41–46] Methylphosphonate substitutions are almost isosteric with phosphodiester linkages, but unlike phosphodiester linkages are neutral and therefore can be used to identify ionic interactions in protein/substrate complexes without introducing steric clashes with the proteins.[41] Methylphosphonate linkages have been shown to induce local bending in the double-helical DNA axis by the mechanism of asymmetric phosphate charge neutralization. However, the bending angle estimated as 3.5° per methylphosphonate substitution[47] is comparable to the intrinsic sequence-specific DNA bending[48] and thermal flexibility of duplex DNA of ~7° per base-pair estimated from its persistence length.[49] Substitution of a methyl group in place of a non-bridging oxygen in the phosphodiester linkage at a point of electrostatic contact with a protein usually decreases the affinity of substrate binding.[41] This property of methylphosphonate modifications makes unnecessary, in most cases, the separation of Rp and Sp stereoisomers of chemically introduced methylphosphonate linkages and justifies the use of their racemic mixtures. 2′-*O*-Methyl substitutions replace the 2′ proton in the deoxyribose ring with a bulky *O*-methyl group with two major outcomes for duplex DNA structure. First, 2′-*O*-methyl groups change the conformational preference of ribose from C2′-*endo* to C3′-*endo* sugar puckering, forcing a local transition from *B*-form to *A*-form DNA. Second, they introduce steric clashes at sites of contacts with the proteins.[45]

Here, we introduced a single methylphosphonate substitution into each phosphodiester linkage of the overlap flap DNA substrate to map phosphates interacting with PfuFEN1. Similarly, we introduced 2′-*O*-methyl substitutions to identify steric contacts in the PfuFEN1/DNA complex. Using the three-dimensional structure of PfuFEN1[23] and a modeled structure of the overlap flap substrate, we performed energy minimization (EM) and molecular dynamics (MD) simulations to test two alternative structures of the PfuFEN1/DNA complex. The model consistent with the methylphosphonate data was used to identify candidate amino acid residues contacting phosphates in the substrate. To confirm the predicted interactions, PfuFEN1 variants mutated at these amino acid residues were tested on the methylphosphonate substrates. The confirmed

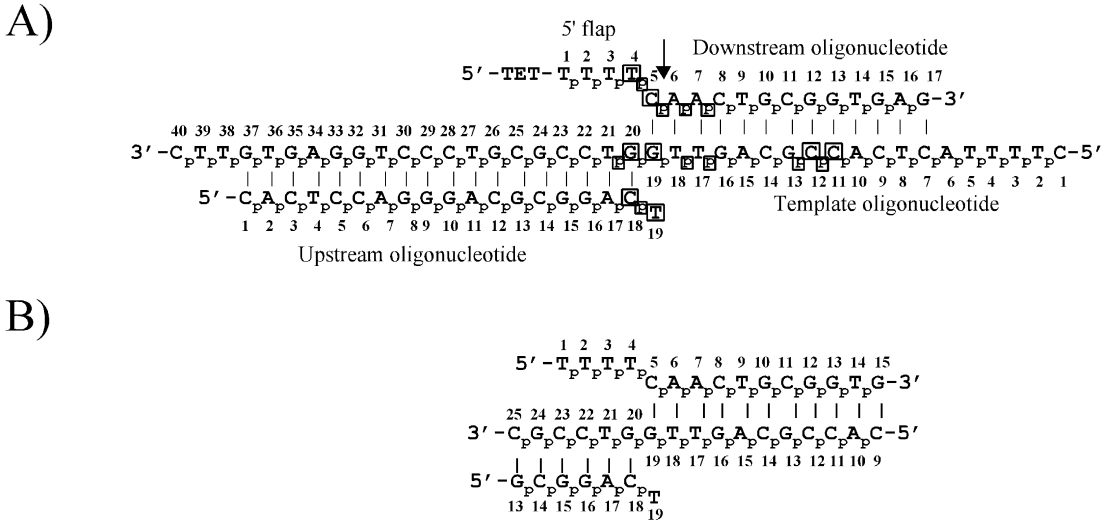

Figure 2. (A) Proposed structure of the overlap flap substrate used in FEN1 enzymatic assays. The substrate includes 19-nt upstream and 17-nt downstream oligonucleotides annealed to a 40-nt template oligonucleotide. Base-pairing is shown by vertical bars. The 5′ flap is the single-stranded part of the downstream oligonucleotide. Nucleosides are numbered from the 5′ end of the oligonucleotides and phosphates are assigned the same number as the adjacent 3′ nucleoside. The specific FEN1 cleavage site is shown by the arrow. Boxed nucleosides and phosphates show positions of 2′-*O*-methyl and methylphosphonate modifications, respectively, which affect the PfuFEN1 cleavage rate. TET is tetracholorofluorecsein dye. (B) Structure of the overlap flap substrate used in MD simulations. The structure was designed using the PfuFEN1 footprint identified in (A). Nucleoside and phosphate residues have the same notation as the corresponding residues in (A).

interactions were used as restraints in MD simulations to develop a detailed model of the PfuFEN1/DNA complex.

Results

Substrate for enzyme activity assays

Figure 2(A) shows the overlap flap substrate designed for PfuFEN1 activity assays. The substrate consists of upstream and downstream oligonucleotides annealed to a template oligonucleotide. The 3′ end nucleotide of the upstream oligonucleotide, a thymine in this substrate, overlaps with the first base-pair, G-C, of the downstream duplex, thus creating the optimal substrate for the structure-specific 5′ nucleases.[8] Although the overlapping nucleotide is important for enzyme activity, it does not need to be complementary to the template and can be mismatched.[21] The downstream oligonucleotide consists of two regions: a 5′ flap region that is non-complementary to the template, and a complementary template-specific region. The substrate is cleaved by the structure-specific 5′ nucleases after the first base-pair of the downstream oligonucleotide, releasing the 5′ flap with one nucleotide of the template-specific region.

The following nomenclature was used to describe the positions of sugar and phosphate residues in the substrate. Each nucleoside is denoted by the corresponding base; a *t*, *u* or *d* subscript is used to indicate template, upstream, or downstream oligonucleotide, respectively; and a number is used to designate the position of this nucleoside relative to the 5′ end of the corresponding oligonucleotide. The phosphates are referred to as P and carry the same subscript and number as the adjacent 3′ nucleoside. For example, the first base-paired nucleoside C at the 5′ end of the downstream oligonucleotide is referred to as C_d5 and the cleavable phosphodiester linkage as P_d6 (see Figure 2(A) for reference).

Enzyme activity assay on modified substrates

Methylphosphonate walking experiments

Methylphosphonate substitutions were used to probe ionic interactions between the PfuFEN1 enzyme and phosphates in the overlap flap substrate. The non-modified (natural) substrate and substrates modified with a single methylphosphonate at each position of the template, upstream, or downstream oligonucleotide were cleaved with PfuFEN1, and the cleavage products were analyzed by gel electrophoresis to determine the initial cleavage rates as described in Materials and Methods (Figure 3). The 50 nM substrate concentration used in these experiments was at least four-fold lower than the PfuFEN1 K_M value of 200 nM for the natural substrate (Table 1). Since the K_M for the natural substrate is expected to have the lowest value among all the substrates, the k_{cat}/K_M values for all substrates were determined from the initial cleavage rates assuming that the substrate concentration is much lower than K_M. The k_{cat}/K_M values normalized for the natural substrate are shown in

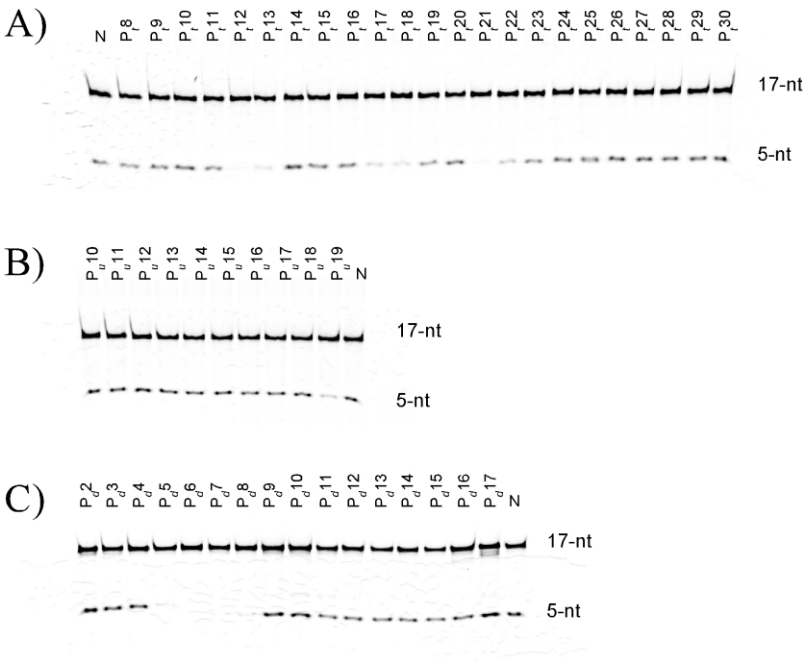

Figure 3. Analysis of PfuFEN1 cleavage products of the natural overlap flap substrate (N) and substrates with methylphosphonate substitutions at the template strand (A), upstream strand (B), or downstream strand (C) by gel electrophoresis. The uncleaved downstream oligonucleotide (17-nt) and cleaved 5′ flap (5-nt) are indicated. Positions of methylphosphonate modification are shown at the top of each panel. Reactions were carried with 0.1 nM PfuFEN1 and 50 nM substrate at 51 °C for five minutes. For brevity, modifications distant from the overlap site that have no effect on cleavage activity are not shown.

Figure 4. Because of ~25% variability of k_{cat}/K_M, only methylphosphonate substitutions that decreased the relative k_{cat}/K_M values by 25% or more were considered to affect the enzyme/substrate interactions. Using these criteria, we hypothesized that positions P_d5–P_d8 and P_u19 in the downstream and upstream oligonucleotides, respectively, were involved in ionic interactions with PfuFEN1. In the template strand, positions P_t12–P_t13, P_t17–P_t18 and P_t21 were identified as three areas of interaction with PfuFEN1. Similar methylphosphonate modification results were obtained with the overlap flap substrate for AfuFEN1 and MjaFEN1 (data not shown).

2′-O-Methyl walking experiments

The substitution of 2′-*O*-methyl at the 2′ position of deoxyribose should introduce steric clashes at sites of close contact with PfuFEN1. Similar to the methylphosphonate walking experiments, we used PfuFEN1 to cleave substrates in which the template, upstream or downstream oligonucleotide contained a single 2′-*O*-methyl substitution at each deoxyribose to determine the initial cleavage rates and calculate k_{cat}/K_M values (Figure 5). Significant decreases in k_{cat}/K_M were demonstrated for 2′-*O*-methyl substitutions at positions T_d4 and C_d5 in the downstream oligonucleotide, at the first and second nucleosides from the 3′ end of the upstream oligonucleotide, T_u19 and C_u18, and at two regions, C_t11–C_t12 and G_t19–G_t20, in the template oligonucleotide. Overall, the 2′-*O*-methyl substitutions identified the same regions of interaction with PfuFEN1 as did the methylphosphonate walking experiments (compare Figures 4 and 5). Particularly, both 2′-*O*-methyl substitutions at C_t11–C_t12 and methylphosphonate substitutions at P_t12–P_t13 inhibited cleavage by PfuFEN1, thus supporting the conclusion that the template strand in

Table 1. Kinetic parameters of the wild-type and mutant PfuFEN1 proteins on the methylphosphonate-modified substrates

		DNA substrate[a]				
Enzyme		Natural DNA	P_t12	P_t21	P_d5	P_u19
PfuFEN1	K_M (μM)	0.20 ± 0.07	1.02 ± 0.49	3.59 ± 1.09	1.61 ± 0.18	1.49 ± 0.22
	k_{cat} (min^{-1})	62.1 ± 22.3	1.8 ± 0.8	7.6 ± 2.1	7.1 ± 0.8	18.3 ± 1.6
R64A	K_M (μM)	0.43 ± 0.06	1.04 ± 0.90	0.38 ± 0.05	1.89 ± 0.73	1.48 ± 0.50
	k_{cat} (min^{-1})	21.8 ± 2.3	$2.7 \pm 2.3 \times 10^{-3}$	7.4 ± 0.9	0.5 ± 0.2	$10.5 \pm 3.0 \times 10^{-3}$
R94A	K_M (μM)	0.21 ± 0.03	0.69 ± 0.38	2.80 ± 0.32	0.25 ± 0.09	1.46 ± 0.58
	k_{cat} (min^{-1})	$12.0 \pm 1.0 \times 10^{-3}$	$1.1 \pm 0.6 \times 10^{-3}$	$3.4 \pm 0.4 \times 10^{-3}$	$24.0 \pm 7.0 \times 10^{-3}$	$9.5 \pm 3.7 \times 10^{-3}$
PfuM3	K_M (μM)	0.90 ± 0.35	7.65 ± 2.60	2.89 ± 1.44	12.95 ± 7.29	8.21 ± 1.27
	k_{cat} (min^{-1})	5.1 ± 1.9	0.1 ± 0.1	1.0 ± 0.5	$5.5 \pm 3.0 \times 10^{-3}$	2.1 ± 0.3

K_M and k_{cat} parameters were determined in 50 mM KCl, 7.5 mM $MgCl_2$, 10 mM Mops at 51 °C.
[a] Position of the methylphosphonate modified phosphodiester linkages.

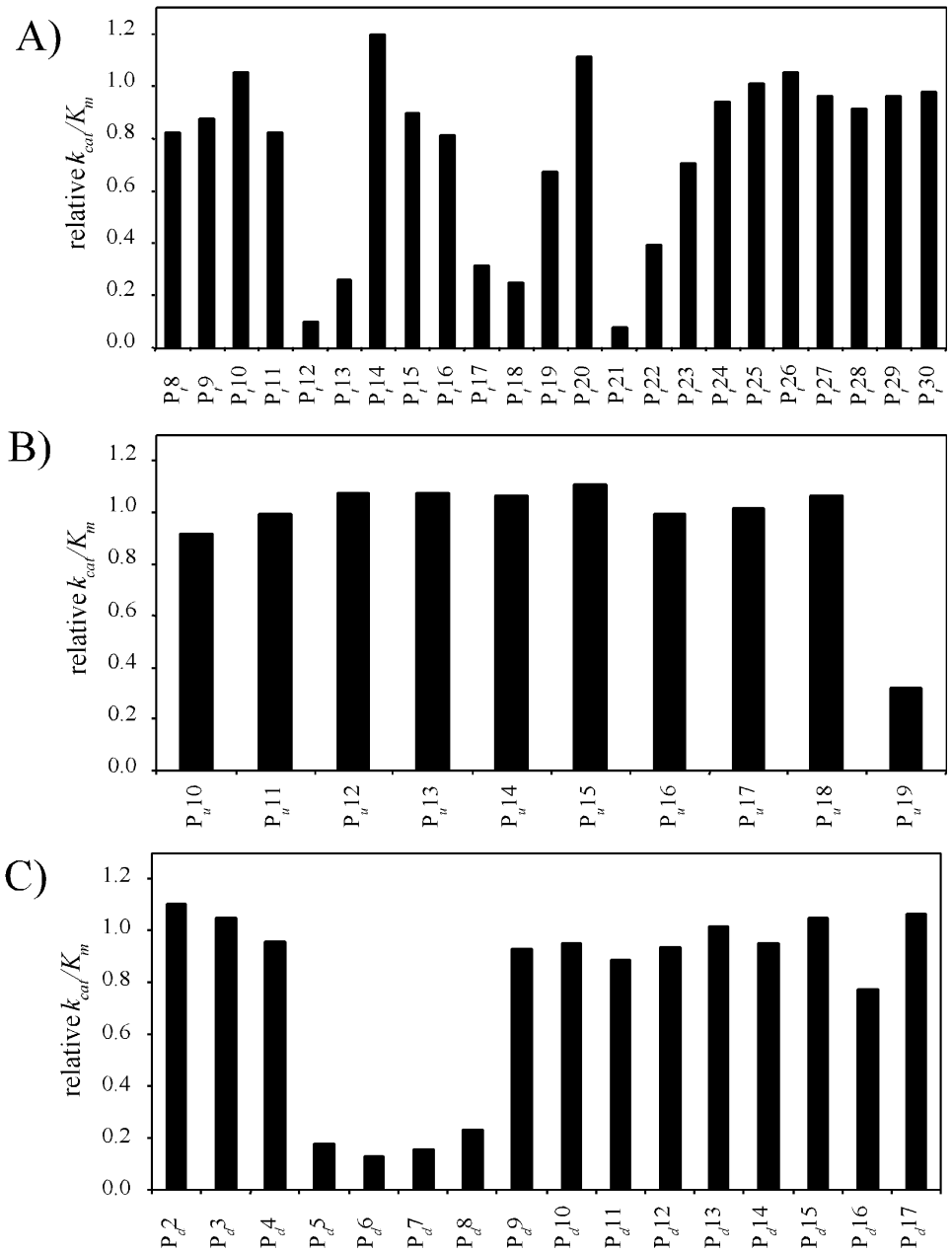

Figure 4. k_{cat}/K_M for PfuFEN1 on substrates with methylphosphonate substitutions at the template strand (A), upstream strand (B), or downstream strand (C) normalized to k_{cat}/K_M of the natural substrate. Results are shown only for modifications closest to the overlap site. The k_{cat}/K_M values were determined from a single point measurement and the coefficient of variation of 25% for the relative k_{cat}/K_M values was estimated from four independent experiments for the natural substrate.

downstream duplex contacts PfuFEN1 eight to nine base-pairs from the overlap.

Orientation of the DNA substrate in PfuFEN1/DNA complex

The first goal of MD simulations was to test two possible orientations of the substrate in the DNA-binding groove of PfuFEN1 and to select the orientation consistent with the methylphosphonate and 2′-*O*-methyl walking experiments. In orientation I, the downstream duplex was placed in the DNA-binding groove close to the HhH motif of PfuFEN1 (residues 223–256) as originally suggested by Kaiser *et al*.[8] In orientation II, the substrate was rotated 180° and placed in the opposite direction relative to the downstream and upstream duplexes, similar to the model proposed for 5′–3′ exonuclease from bacteriophage T5.[26] For both orientations, the minimal restraints imposed on the enzyme–DNA interactions defined the distances between the magnesium ion M-1 and the cleavable phosphodiester bond P_d6 and between the metal ions and the active site of the enzyme.

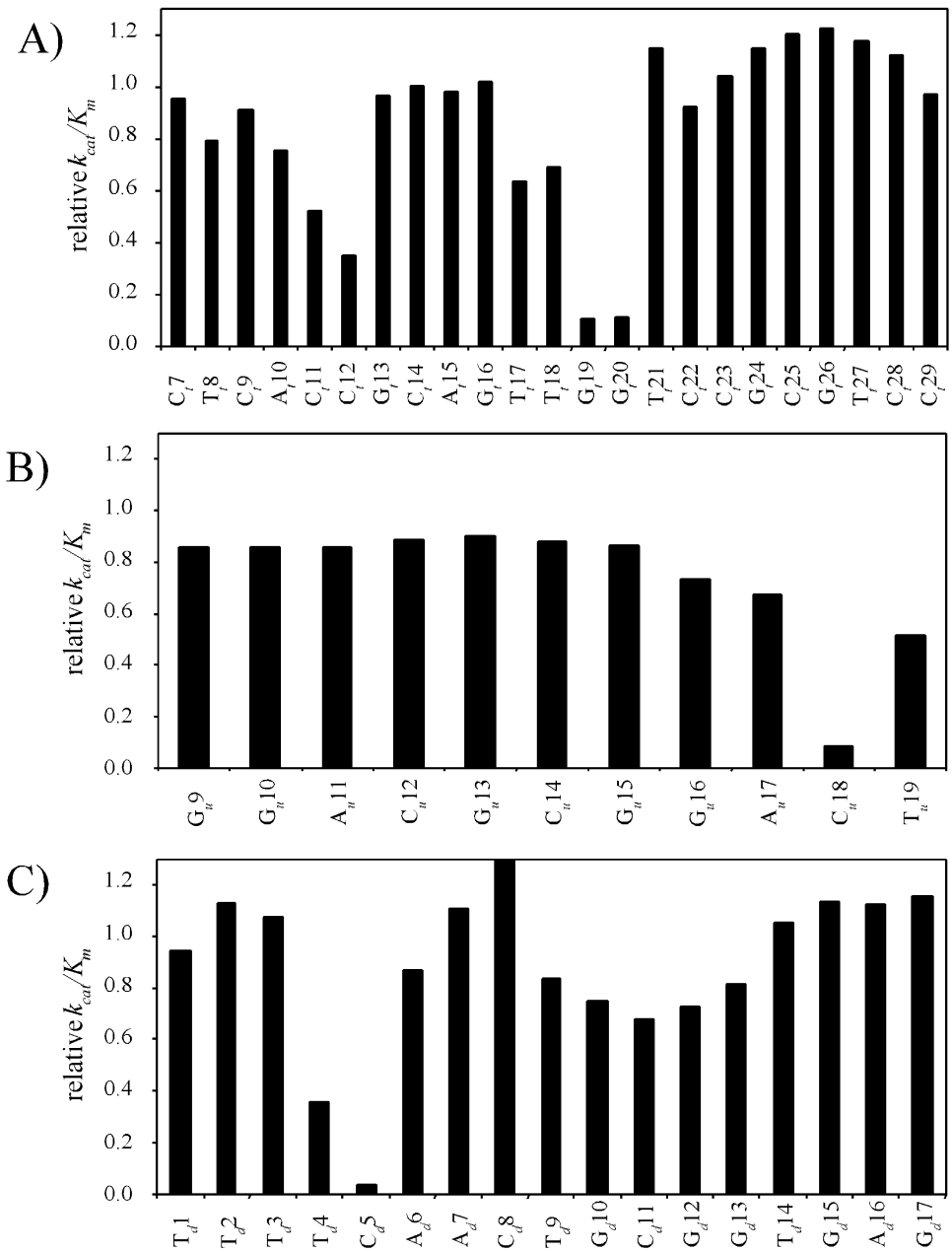

Figure 5. k_{cat}/K_M values for PfuFEN1 on substrates with 2′-*O*-methyl substitutions at the template strand (A), upstream strand (B), or downstream strand (C) normalized to k_{cat}/K_M of the natural substrate. Results are shown only for modifications closest to the overlap site. The k_{cat}/K_M values were determined from a single point measurement and the coefficient of variation of 25% for the relative k_{cat}/K_M values was estimated from four independent experiments for the natural substrate.

Table 2 summarizes the minimal restraints used in the MD calculations.

MD simulations predict different structures for orientations I and II (Figure 6). In particular, in orientation I, the template strand (green) in the downstream duplex interacts with the HhH motif (red) in the C_t11–C_t14 region six to nine base-pairs from the overlap. On the other hand, in orientation II, PfuFEN1 enzyme contacts with the template strand in the downstream duplex are confined to the G_t16–G_t19 region. The interactions predicted in orientation I, but not those in orientation II, are consistent with the results of enzyme activity assays on the methylphosphonate and 2′-*O*-methyl-modified substrates, which suggest that the P_t12–P_t13 phosphates and C_t11–C_t12 deoxyribose residues are involved in contacts with PfuFEN1.

Mutational analysis of interactions in the PfuFEN1/DNA complex

To provide additional support for orientation I and to identify specific PfuFEN1/DNA

Table 2. Distance restraints used in MD calculations of PfuFEN1/DNA complex

Interaction[a]		Distance restraint (Å)	Reference
		I Minimal restraints[b]	
M-1	P_d6	3.5 ± 1.5	Cleaved phosphodiester bond and active site restraint
M-1	Asp27	5.0 ± 2.5	23
M-1	Asp80		
M-1	Glu152		
M-1	Glu154		
M-2	Asp173	5.0 ± 2.5	23
M-2	Asp175		
M-2	Asp236		
M-1	M-2	5.5 ± 2.0	23
Watson–Crick hydrogen bonding restraints		2.2 ± 0.5	This work
		II Residue-specific restraints[b]	
Arg94	P_d5	5.5 ± 2.5	This work
Arg94	P_d6	5.5 ± 2.5	
Arg64	P_t21	5.5 ± 2.5	This work
P_t12	Gly246	3.5 ± 1.0	50,51
P_t12	Gly247		
P_t12	Gly248		
P_t13	Gly244	3.5 ± 1.0	50,51

[a] The center of mass of the carboxy groups of Asp and Glu or the amino groups of Arg side-chains were used to define distance restraints involving those residues. For G246, G247, G248, or G244, restraints were defined between the backbone nitrogen atoms of those residues and the center of mass of the phosphorous groups of P_t12 or P_t13 and were based on the crystal structure of the human polymerase β[50] and *E. coli* DNA glycosylase II.[51]

[b] Minimal restraints were initially used for both orientations I and II of the PfuFEN1/DNA complex. Residue-specific restraints were used in combination with minimal restraints to refine the structure of orientation I (see Materials and Methods).

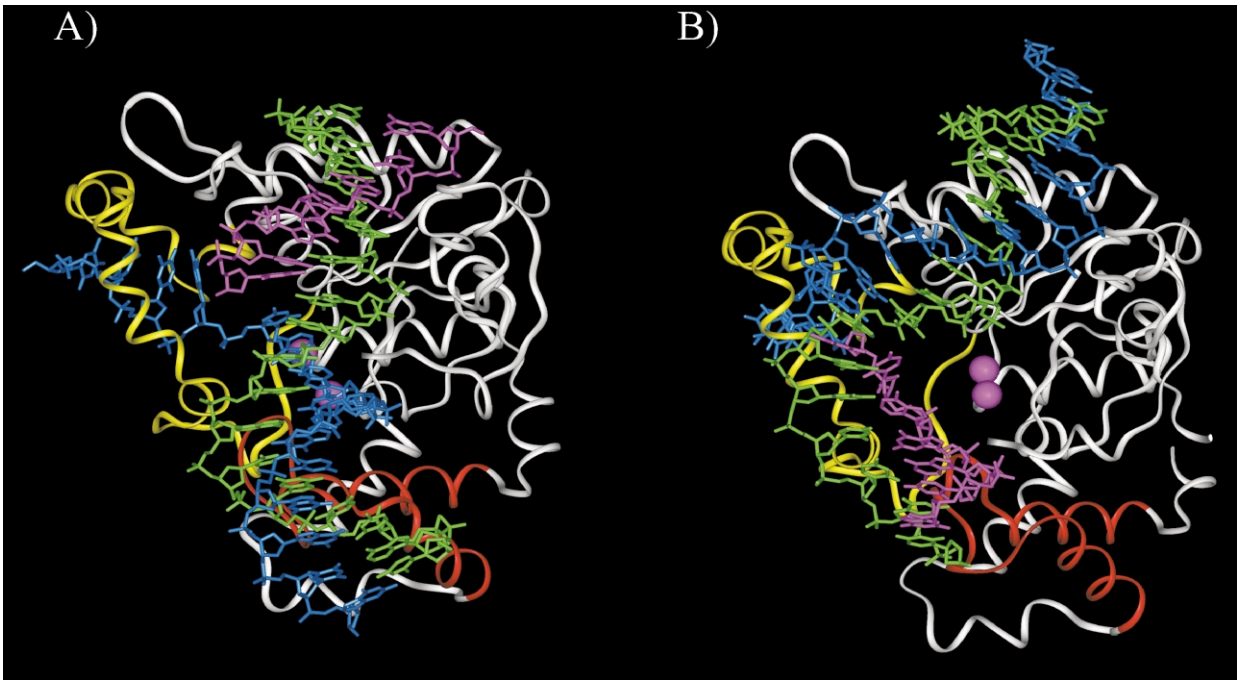

Figure 6. Modeled 1 ns structures of PfuFEN1/DNA complexes in orientations I (A) and II (B). For both orientations, the cleaved phosphodiester bond P_d6 of the DNA was restrained at distance of 2.5 Å to 4.5 Å from the M-1 magnesium ion. The template, upstream and downstream oligonucleotides are shown in green, magenta, and blue, respectively. In orientation I, the DNA is located in the DNA-binding groove with the downstream duplex (green–blue) facing the HhH motif (red) of the enzyme. In orientation II, the DNA was rotated 180° relative to the position of the downstream and upstream duplexes in the DNA-binding groove.

interactions, amino acids from four different regions in the DNA-binding groove were selected for mutational analysis. Group 1 includes the Y33, Q34, R40, and R64 amino acid residues located in the α2 and α3 helixes of PfuFEN1[23] that presumably interact with the upstream duplex in the orientation I model. Group 2 includes the R94, R95, and R98 amino acid residues from the C terminus of the helical arch region, which are predicted to interact with P_d5 of the probe strand. Group 3 includes all positively charged amino acids K193, R194, K195, K199, and K206 in the PfuFEN1 β_A/β_B hairpin[23] and the Q172 and R186 amino acids, which are predicted to interact with the G_u15–C_u18 region of the upstream strand. Also, this group included the PfuM3 mutant in which the β_A/β_B hairpin was replaced with a shorter loop of amino acids Lys-Glu-Met from MjaFEN1 (residues 192–194). This loop was identified as a counterpart of the β_A/β_B hairpin in a structural alignment of PfuFEN1 and MjaFEN1. Group 4 includes the K243, K248, and K249 amino acids of the HhH motif and amino acids K266 and Q267 of the α12/α13 hairpin (261–280) predicted to interact with P_t11–P_t13 of the template strand.

Individual amino acid residues from all four groups were substituted with alanine to produce PfuFEN1 point mutants. We assumed that a methylphosphonate substitution at a phosphate forming an ionic contact with a positively charged amino acid would have a smaller effect on the relative activity of the PfuFEN1 variant if this amino acid was changed to an uncharged amino acid, e.g. alanine. Therefore, the PfuFEN1 point mutants were tested in the activity assay on methylphosphonate-modified substrates, and relative k_{cat}/K_M values obtained for each mutant were compared to those obtained using the wild-type PfuFEN1 on the methylphosphonate-modified substrates.

Among group 1 mutants, the Y33A mutation produced an inactive enzyme, and both the Q34A and R40A mutations showed no significant effect on relative k_{cat}/K_M for all methylphosphonate-modified substrates (data not shown). However, the R64A mutant produced a k_{cat}/K_M profile on the template strand that was different from that of wild-type PfuFEN1 (compare Figures 7(A) and 4(A)). A significant increase in relative k_{cat}/K_M was observed for the R64A mutant on the P_t21 methylphosphonate substrate. Interestingly, the R64A mutation decreased relative k_{cat}/K_M values for the P_t11, P_t23 and P_t24 methylphosphonate substrates.

In group 2, the R95A mutant showed no activity, and the R98A mutant showed no effect on the k_{cat}/K_M profiles on all methylphosphonate DNA substrates (data not shown). The R94A mutation had no effect on the k_{cat}/K_M profiles on the template and upstream strands. However, in the downstream strand, a significant increase in relative k_{cat}/K_M was observed on the P_d5 and P_d6 methylphosphonate substrates (compare Figures 7(B) and 4(C)). The increase in k_{cat}/K_M of the R94A mutant on the P_d6 methylphosphonate substrate is surprising since P_d6 is the phosphodiester linkage cleaved by PfuFEN1, and its modification with a methylphosphonate is expected to have a strong inhibitory effect on enzyme activity. The P_d6 methylphosphonate substrate cleaved by the R94A mutant produced a 5′ end labeled product of the same size as the unmodified substrate, demonstrating that the cleavage took place at the P_d6 methylphosphonate linkage. However, prolonged treatment of the P_d6 methylphosphonate substrate with the R94A mutant showed only approximately 50% substrate cleavage, suggesting that only one of two stereoisomers was cleaved by the mutant.

None of the Q172A, R186A, K193A, R194A, K195A, K199A, or K206A point mutants included in group 3 showed a significant effect on the k_{cat}/K_M profile of any of the methylphosphonate substrates (data not shown). The PfuM3 mutant demonstrated an increase in relative k_{cat}/K_M on the P_t17, P_t18, P_t21, and P_t22 methylphosphonate substrates compared to the wild-type PfuFEN1 (Figure 7(C)). Finally, no differences in the k_{cat}/K_M profiles on any of the methylphosphonate substrates were observed for the group 4 point mutants K243A, K248A, K249A, K266A, Q267A and the K248A/K249A double mutant.

To confirm the k_{cat}/K_M results, individual K_M and k_{cat} values for the R64A, R94A, PfuM3 mutants, and wild-type PfuFEN1 on the P_t12, P_t21, P_d5, and P_t19 methylphosphonate substrates and the natural substrate were determined as described in Materials and Methods (Table 1). Comparison of the wild-type PfuFEN1 and the R64A mutant shows that the P_t21 methylphosphonate substitution increases the K_M value of PfuFEN1 approximately 18-fold but has virtually no effect on the K_M of the R64A mutant. The P_d5 methylphosphonate substitution does not practically affect the kinetic parameters for the R94A mutant; however, it increases the wild-type PfuFEN1 K_M value approximately eightfold and decreases the K_{cat} value ninefold. Finally, the P_t21 modification increases the K_M value for PfuM3 threefold compared to an approximately 18-fold increase for PfuFEN1. A similar decrease in k_{cat} is observed for both enzymes. Overall, the individual K_M and k_{cat} values shown in Table 1 agree with the k_{cat}/K_M results obtained in the enzyme activity assays.

In conclusion, the mutational analysis supports orientation I of the PfuFEN1/DNA complex and suggests that R94 and R64 contact P_d5 and P_t21, respectively, and that the β_A/β_B loop contacts the P_t17–P_t18 and P_t21–P_t22 regions in the template strand.

MD simulations of the PfuFEN1/DNA complex with residue-specific restraints

The interactions revealed by the mutational analysis were used to improve the structural

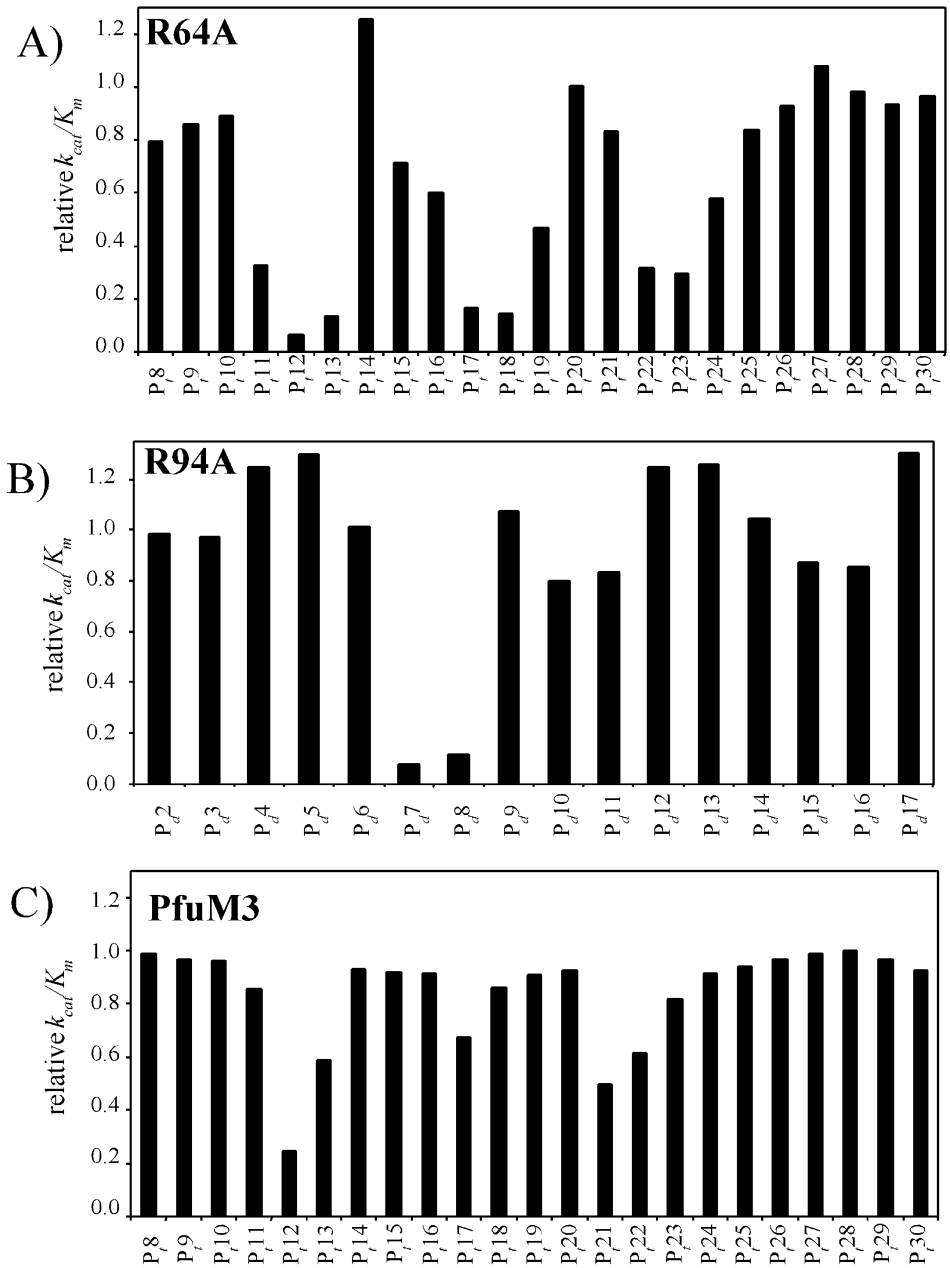

Figure 7. Relative k_{cat}/K_M values for R64A (A), R94A (B) and PfuM3 (C) mutants on the substrates with methylphosphonate substitutions. The k_{cat}/K_M values were normalized for that of the natural substrate. Results are shown only for regions where changes in k_{cat}/K_M profile for each of the mutants occurred. Reactions were performed with 50 nM substrate using 0.5 nM enzyme at 51 °C for ten minutes. The coefficient of variation of the relative k_{cat}/K_M values determined from four independent experiments with the natural substrate was 25%.

model of the PfuFEN1/DNA complex by introducing distance restraints between specific residues of the enzyme and the DNA into MD simulations. In the minimal restraints model (Figure 6(A)), the distance between the center of mass the amino group of R64 and the phosphorous in P_t21 is 17.6 Å and those for the R94–P_d5 and R94–P_d6 interactions are 11.7 Å and 14.9 Å, respectively. Based on the proposed R64A–P_t21 and R94A–P_d5 contacts, we introduced a distance restraint of <8 Å between the center of mass of the amino groups of the side-chain of R64 and R94 and the phosphorous atom of P_t21 and P_d5, respectively.

Although the mutational analysis did not predict interactions between the positively charged amino acid residues of the PfuFEN1 HhH motif and the DNA, it was assumed that the HhH motif contacts the P_t13 and P_t12 phosphates in the template strand based on the methylphosphonate walking experiments (Figure 4(A)). To define restraints for the HhH motif, we used co-crystal structures for human polymerase β (PDB accession 1BPY)[50] and *Escherichia coli* 3-methyladenine DNA glycosylase

II (1DIZ)[51] that demonstrate interactions of the cognate HhH motifs with double-stranded DNA substrates. The HhH interactions in these structures occur through hydrogen bonds of <3.3 Å between oxygen atoms of two adjacent DNA phosphates and the backbone nitrogen atoms of the G105, G107, 109, and A110 residues in polymerase β and the G214, G216, and T219 residues in glycosylase. The corresponding residues in the PfuFEN1 HhH motif, G244, G246, K248, and K249, were identified by structure-based sequence alignment.[32] In the minimal restraints model, the distances between the backbone nitrogen atoms of G244, G246, K248, and K249 and the nearest oxygen of P_t13 or P_t12 are 10.9, 15.6, 18.2, and 18.2 Å, respectively. To introduce hydrogen bond interactions observed in the co-crystals, the distances between the G244, G246, K248, and K249 backbone nitrogen atoms and a center of mass of phosphorus and two non-bridging oxygen atoms of P_t13 and P_t12 were restrained to $3.5(\pm 1.0)$ Å. Table 2 summarizes the residue-specific restraints used in the MD calculations.

Figure 8 shows a 5 ns MD structure of the PfuFEN1/DNA complex using residue-specific restraints. In the residue-specific model, the substrate moved deeper in the active-site DNA-binding groove compared with the minimal restraints model. For instance, the distance between the center of mass of the arginine amino groups and phosphorous in R64–P_t21, R94–P_d5, and R94–P_d6 pairs is decreased to 3.8, 3.8, and 4.3 Å, respectively. The DNA also moved closer to the HhH motif to form hydrogen bonds O2–P_t13–G244, O2–P_t12–G246, O1–P_t12–K248, and O1–P_t12–K249 of 2.77, 3.7, 3.3, and 3.0 Å, respectively. In addition, positively charged residues of the HhH motif demonstrate interactions with P_d8 in the downstream strand and with the major groove of the downstream duplex in the C_t11–C_t13 region. The upstream and downstream duplexes are in the *B*-form conformation and form an angle of approximately 40° between duplex axes at the overlap. The M-1 magnesium ion interacts with the cleavable phosphodiester bond P_d6 and the M-2 ion, which is not bound by any restraints, interacts with the P_d6 and P_d7 phosphates. Interactions predicted by the final structure are summarized in Figure 9.

Discussion

Here, we have modeled a three-dimensional structure of the PfuFEN1 structure-specific 5′ nuclease in its complex with DNA using MD simulations guided by enzyme–DNA distance restraints derived from experimental data. Regions of the overlap flap substrate involved in specific contacts with PfuFEN1 were identified using methylphosphonate and 2′-*O*-methyl walking experiments. Most contacts were observed close to the cleavage site, however, an additional region of both electrostatic and steric interactions was identified on the template oligonucleotide at positions P_t12, P_t13, C_t11, and C_t12 (Figure 2(A)). The asymmetry of the footprint, which extends nine base-pairs into the downstream duplex and only two base-pairs into the upstream duplex, agrees well with analyses of enzyme activity on overlap flap substrates of different length[8] and phosphate ethylation interference experiments with a flap substrate and the 5′ nuclease domain of DNA polymerase I from *E. coli*.[52]

The PfuFEN1 footprint was used to determine the orientation of the substrate in the DNA-binding groove. Originally proposed models for the 5′ nuclease/DNA complex suggested that the upstream duplex of the flap substrate interacts with the HhH motif of the 5′ nuclease enzymes.[7,26,53] However, based on the minimal substrate length requirements, Kaiser *et al*.[8] proposed that the HhH motif interacts with the downstream duplex of the overlap flap substrate. A similar model was also proposed by Dervan *et al*.[40] from kinetic and binding characteristics of T5 5′ nuclease mutants using hairpin-like substrates containing only a downstream duplex. MD simulations using only minimal restraints specifying distance between the cleavable linkage P_d6 and M-1 magnesium ion showed that only orientation I, in which the downstream duplex interacts with the HhH motif, was consistent with the asymmetric pattern of the PfuFEN1 footprint and was proposed as a minimal restraints model of the PfuFEN1/DNA complex (Figure 6(A)).

Binding of DNA substrates often induces significant conformational changes in protein structure. For example, comparison of the apo and DNA complex structures of the Klenow fragment of *E. coli* DNA polymerase I demonstrates that the thumb region of the protein moves as much as 12 Å to accommodate the DNA.[54] There is no indication of induced conformational changes in the minimal restraints model, and it is unlikely that such changes, which may occur at a microsecond or even millisecond time range,[55] would be observed in nanosecond-long MD models. The importance of these considerations can be illustrated by the nature of the interactions between the HhH motif and the downstream duplex predicted by the minimal restraints model to occur through ionic contacts between the side-chains of K248 and K249 and the P_t12–P_t13 phosphates. These interactions contradict the co-crystal structures of human polymerase β[50] and *E. coli* 3-methyladenine DNA glycosylase II,[51] demonstrating that the HhH motif interacts with double-stranded DNA by forming hydrogen bonds between the backbone nitrogen atoms of conserved HhH residues and oxygen atoms of two adjacent DNA phosphates. The contacts suggested from the co-crystal structures would require that the downstream duplex in the minimal restraints model move closer to the PfuFEN1 HhH motif and form specific hydrogen bonds with

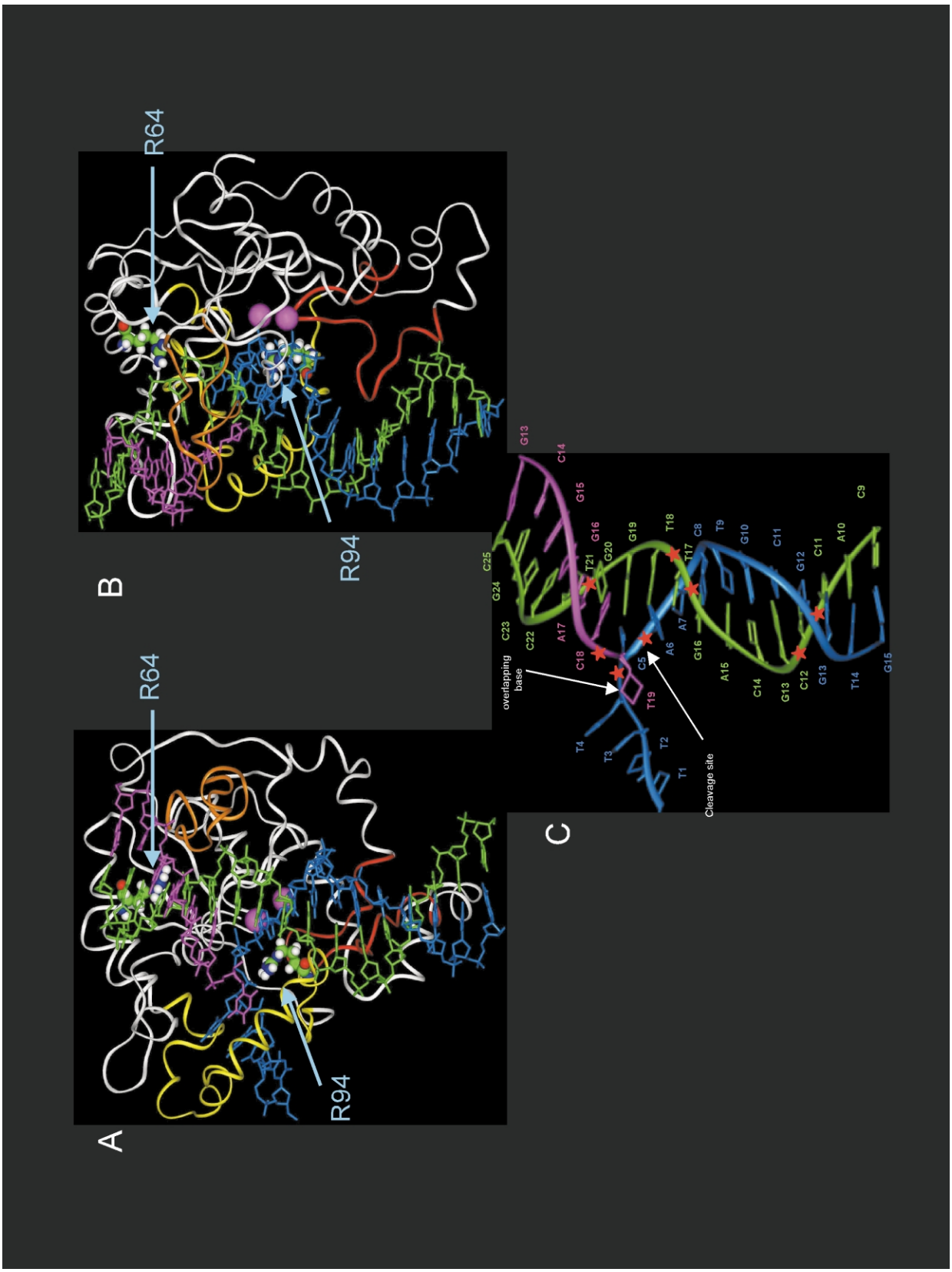

Figure 8 (*legend opposite*)

backbone nitrogen atoms of G244, G246, K248, and K249.

To experimentally determine specific interactions in the PfuFEN1/DNA complex, we used enzyme activity assays with mutant PfuFEN1 enzymes on methylphosphonate-containing substrates. The basis for this approach has been previously proven for specific steric interactions between the MS2 bacteriophage coat protein and its cognate RNA hairpin.[56] The methylphosphonate profiling of the PfuFEN1 mutants identified specific ionic contacts between R94 and P_d5 and between R64 and P_t21. A similar approach was previously used to identify specific contacts between the 5′ nuclease domain of DNA polymerase I from *E. coli* and a flap substrate[52] with methylphosphonate substitutions studied in the downstream strand at positions corresponding to P_d3, P_d4, P_d5 and P_d7. The authors observed a similar inhibitory effect on the cleavage rate of substrate modified at positions analogous to P_d5 and P_d7 and, based on mutational analysis, suggested that these phosphates interact with Y77, K78, and R81 and R20 of the *E. coli* 5′ nuclease domain, respectively. Alignment of the *E. coli* 5′ nuclease domains and TaqExo sequences followed by three-dimensional structure alignment of TaqExo and PfuFEN1 suggests that Y77, K78, R81, and R20 of the *E. coli* 5′ nuclease domain correspond to R94, R95, R98, and Q34 of PfuFEN1, respectively. Here we show that R94 contacts P_d5, but the R95A mutation results in an inactive enzyme and both the Q34A and R98A mutations have no significant effect on the methylphosphonate profiling. These differences can be explained by a misalignment of the *E. coli* 5′ nuclease domain and PfuFEN1 due to low sequence similarity and differences in the binding characteristics of the overlap flap and flap substrates.

Introduction of the residue-specific restraints moved the substrate deeper in the DNA-binding groove compared to the minimal restraints model. Superposition of enzyme structures in the minimal and final models shows that the P_t13 and P_t12 phosphates in the downstream duplex move by ~20 Å and the upstream duplex moves by 17 Å on average. The tighter contacts between the DNA and enzyme became possible because of conformational changes in the enzyme structure: the C^α atoms of the β_A/β_B loop (192–206) moved by 11 Å on average with a maximum distance of 16.5 Å observed for P197. The final PfuFEN1/DNA model was challenged to explain differences between the methylphosphonate and 2′-*O*-methyl PfuFEN1 footprints. The latter identifies steric interactions that have not been used to specify the distance restraints in MD simulations. In the downstream strand, these differences are manifested by an effect of the P_d6, P_d7, and P_d8 methylphosphonate substitutions on k_{cat}/K_M and the lack of such an effect for the A_d6 and A_d7 2′-*O*-methyl substitutions (Figure 2(A)). The modeled structure explains this pattern by the existence of contacts between P_d6–P_d8 and the magnesium ions and K248 whereas the A_d6 and A_d7 nucleosides are exposed in solution outside the DNA-binding groove and do not contact PfuFEN1. Conversely, the G_t20 and G_t19 2′-*O*-methyl modifications, but not the P_t20 and P_t19 methylphosphonate modifications, affect k_{cat}/K_M. Again, the model predicts that the G_t20 and G_t19 nucleosides face the DNA-binding groove, and the 2′-*O*-methyl substitutions at these positions are likely to cause a steric clash with PfuFEN1. To the contrary, no specific interactions are predicted for P_t20 and P_t19.

There are several interactions suggested from the PfuFEN1 footprints that cannot be explained by the final model. For example, methylphosphonate substitutions at P_t17 and P_t18 decrease k_{cat}/K_M (Figure 4(A)); however, no contacts involving these phosphates are predicted in the model. Potential candidates for these contacts include a group of positively charged lysine residues 87, 89, 93 and arginine residues 94, 95, 98, 106 in the helical arch region; however, the methylphosphonate profiling does not support this hypothesis for the tested amino acid residues, R94 and R98, and the R95A mutation has produced an inactive enzyme. Interaction between the helical arch and the P_t17 and P_t18 phosphates might be only possible as a result of large substrate-induced conformational changes in the helical arch, such as previously suggested for human FEN1 from Fourier transform infrared spectroscopy demonstrating an increase in helicity of the helical arch region upon DNA binding.[57] Some interactions suggested in the model and supported by PfuFEN1 footprint data still need to be confirmed by mutational analysis. For example, according to the model, P_u19 forms ion pairs of 2.9 Å, 3.0 Å, and 3.9 Å with K104, K111, and R40, respectively, and the methylphosphonate substitution at P_u19 decreases k_{cat}/K_M value of PfuFEN1 (Figure 4(B)). Only one of these amino acid residues, R40, was

Figure 8. 5 ns MD model of the PfuFEN1/DNA complex with residue-specific restraints. (A) A view showing the overlap flap substrate bound in the DNA-binding groove of PfuFEN1. The HhH motif, β_A/β_B hairpin, helical arch and magnesium ions are shown in red, orange, yellow and pink, respectively. Arginine residues 94 and 64 are displayed with a space-filled model. The template, upstream and downstream oligonucleotides are shown in green, magenta, and blue, respectively. (B) 90 degree rotated view of the model shown in (A). (C) A cartoon description of the substrate structure in the model (adopted from the view (A)). The 3′ end overlapping base of the upstream oligonucleotide is shown by arrows. The phosphodiester linkages P_d5, P_d6, P_t12, P_t13, P_t17, P_t18, P_t21, and P_u19 are shown with asterisks.

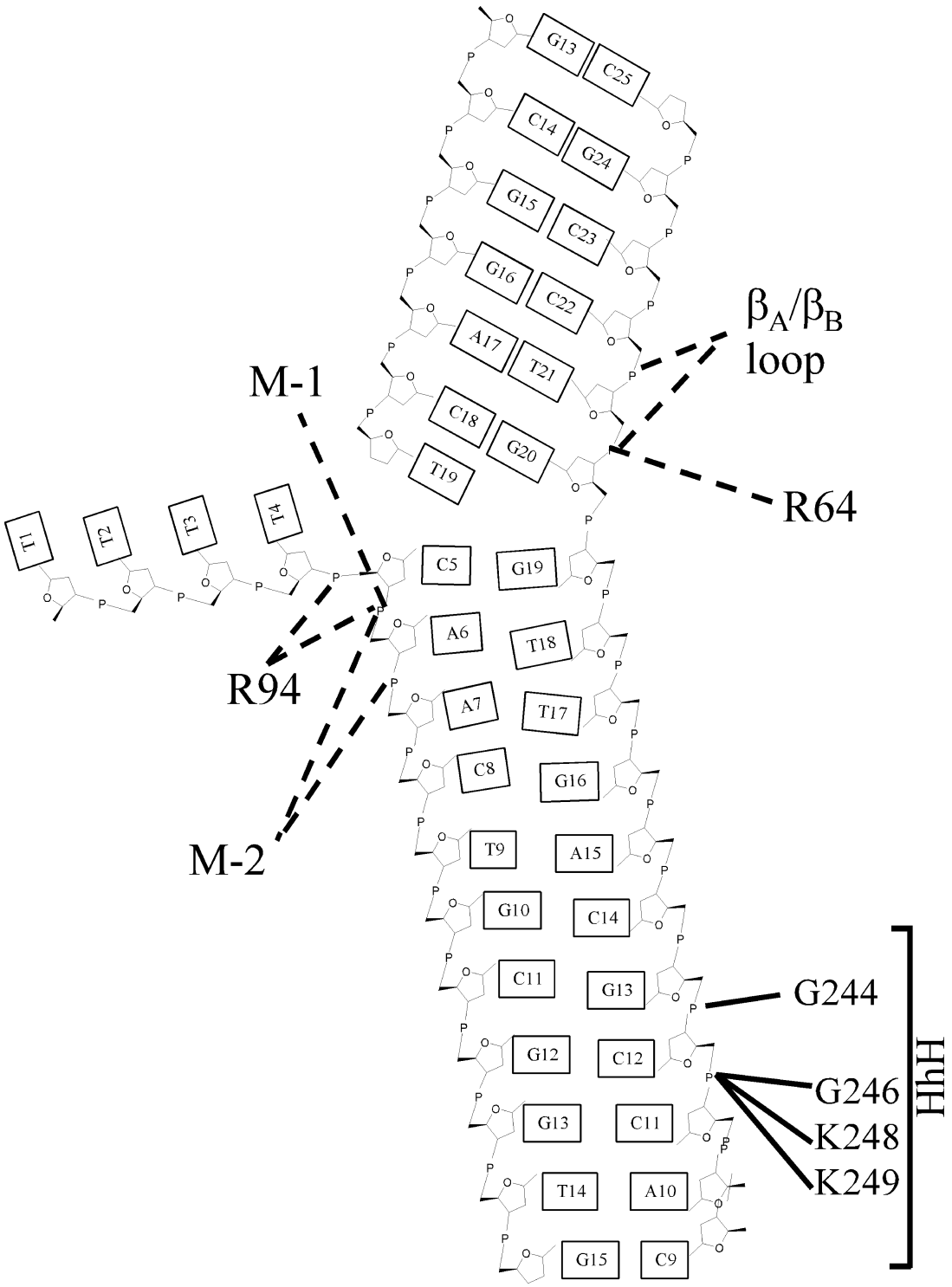

Figure 9. Schematic representation of interactions proposed in the PfuFEN1/DNA complex. Hydrogen bonds and ion pairs are shown by continuous and broken lines, respectively. Backbone nitrogen atoms of the G244, G246, K248, and K249 amino acid residues in the HhH motif form hydrogen bonds with P_t12 and P_t13 in the template oligonucleotide. Amino acid R94 forms ion pairs with P_d5 and P_d6 in the downstream oligonucleotide and amino acid R64 contacts P_t21 in the template oligonucleotide. The β_A/β_B loop interacts with the upstream duplex between residues P_t21 and P_t22.

tested in the methylphosphonate profiling experiments, but showed no interaction with P_u19.

The 3′ terminal nucleotide of the upstream strand is an important element in substrate recognition by FEN1 enzymes. The final model predicts that distance between centers of the 3′ nucleotide and the Y33 amino acid aromatic rings measured over 100 snapshots taken from the last 1 ns of 5 ns

MD simulation is ~6(±2) Å. This distance corresponds to a maximum of distance distribution for aromatic amino acid pairs in X-ray crystal structures of non-homologous proteins[58] suggesting stacking interactions between the 3′ nucleotide and the Y33 residue. In support of this conclusion, Y33 is highly conserved in the FEN1 family, and substitution of this amino acid with alanine or glutamic acid inactivates PfuFEN1, but its substitution with another aromatic amino acid, phenylalanine, reduces PfuFEN1 activity only by 50% (data not shown). The exact position of the 5′ flap, which according to the threading model should be within the helical arch of PfuFEN1,[5,29] is not defined in the final model. During the dynamics the flap moves in and out of the helical arch due to the high flexibility of both the helical arch and the 5′ flap, thus making it difficult to interpret its specific contacts. It is possible that conformational changes suggested for the helical arch should be included in the model to determine specific interactions of the 5′ flap and the 3′ terminal nucleotide of the upstream strand.

Significant progress in understanding the mechanism of lagging strand synthesis in eukaryotes led to a model describing the synthesis as a series of highly coordinated steps performed by the Okazaki fragment processing complex, which includes FEN1.[59,60] It is thought that the synchronized product release in one step and its binding as a substrate in the next step is governed by substrate specificity of the proteins constituting the complex. FEN1 plays a critical role in the transition from the strand displacement step performed by eukaryotic polymerase δ to the 5′ flap removal step and in the passing of the nick substrate to DNA ligase I. The first detailed model of PfuFEN1/DNA complex proposed here is an important step in understanding the organization of Okazaki fragment processing complex and dynamics of complex rearrangement during the lagging strand maturation. Given the consistency of our model with methylphosphonate and 2′-*O*-methyl walking results, we believe that it provides a correct general orientation of the overlap flap substrate in the active-site DNA-binding groove of PfuFEN1 and defines likely contacts in the PfuFEN1/DNA complex. The modeled PfuFEN1/DNA structure may be further improved by incorporating new experimental data.

Materials and Methods

DNA synthesis and purification

DNA oligonucleotides were synthesized with an Expedite 8909 synthesizer (PerSeptive Biosystems) using standard phosphoramidite chemistry. All phosphoramidites including tetracholorofluorecsein (TET) dye, 2′-*O*-methyl, and methylphosphonate phosphoramidites were purchased from Glen Research, Sterling, VA. Synthesis and deprotection were performed according to the vendor's recommended procedures and further purification was performed by electrophoresis on a 20% (w/v) denaturing acrylamide gel as described.[8] Methylphosphonate oligonucleotides were used without separation of Rp and Sp isomers. The purity of all oligonucleotides used in enzyme activity assays was analyzed by capillary electrophoresis on a P/ACE MDQ CE system (Beckman) with μPAGE-5 capillaries (Agilent Technologies, Palo Alto, CA). Oligonucleotide concentrations were determined from the absorption at 260 nm using extinction coefficients for nucleosides and dinucleotide monophosphates.[61]

Mutagenesis

The Y33A, Q34A, R40A, R64A, R94A, R95A, R98A, Q172A, R186A, K193A, R194A, K195A, K199A, K206A, K243A, K248A, K249A, K266A, and Q267A mutants were prepared by site-directed mutagenesis with a Quikchange kit (Stratagene) using 30 cycles of PCR amplification on the pTrc99a-PfuFEN1 plasmid DNA[8] with 200 nM mutagenic primers. After *Dpn*I digestion the amplified DNA was transformed into *E. coli* DH5α by electroporation. The presence of desired mutations was verified by sequencing. The PfuM3 deletion variant was prepared by replacing amino acid residues GKRKLPGK NVYVEIK of PfuFEN1 at position 192–206 with amino acids KEM of MjaFEN1 (192–194) using recombinant PCR. The 5′ half of the PfuM3 gene was amplified with 5′-TGTGGAATTGTGAGCGG (pTrcFwd) and 5′-CTCTG GCATCTCCTTTGTTATTGTTAAGTTTCTAAC primers and the 3′ half was amplified with 5′-ACAAAGGAGAT GCCAGAGTTGATAATTTTGGAGGAAG and 5′-TAAT CTGTATCAGGCTG (pTrcRev) primers. The DNA products were purified from a 1% (w/v) agarose gel using a Qiaquick kit (Qiagen) and used in PCR reaction with the pTrcFwd and pTrcRev primers. The amplified DNA was gel-purified and cloned into pTrc99a vector as described.[8]

Enzyme expression and purification

The plasmids carrying PfuFEN1 variants were transformed into *E. coli* BL21 (Novagen) for expression and purification of the proteins was performed as described.[8] At the final step of purification the proteins were dialyzed and diluted in a buffer containing 50% (v/v) glycerol, 20 mM Tris–HCl (pH 8), 50 mM KCl, 0.5% Tween 20, 0.5% Nonidet P40, 100 μg/mL BSA.

Kinetic parameter measurement and enzyme activity assays

Kinetic parameters, K_M and k_{cat}, for the wild-type and mutant PfuFEN1 enzymes were determined with the substrate shown in Figure 2(A). Reactions were prepared by mixing the downstream, upstream and template oligonucleotides at 5 μM concentration each in a buffer containing 100 mM Mops, 75 mM $MgCl_2$ and heating the sample to 95 °C for one minute and slow cooling to room temperature. The annealed substrate was stored at −20 °C. The kinetics experiments were performed at different substrate concentrations from 10 nM to 800 nM and constant enzyme concentration between 0.1 nM and 100 nM, depending on the specific activity of a particular enzyme. Enzyme/substrate mixtures were prepared in 100 μl 10 mM Mops (pH 7.5), 50 mM KCl, 7.5 mM $MgCl_2$, 10 ng/μl tRNA on ice, and the reactions were

started by incubation at 51 °C. Aliquots (10 μl) of each reaction were then transferred into pre-chilled tubes containing 15 μl 95% formamide, 10 mM EDTA, and 0.02% (w/v) methyl violet blue (Sigma-Aldrich) at time points of 5, 10, 15, 30, and 60 minutes. Aliquots (5 μl) of the stopped reactions were then analyzed by gel electrophoresis to obtain initial rates of the cleavage reactions as described.[8] The K_M and k_{cat} values were calculated from the initial rates using the Michaelis–Menten equation. The enzyme activity assay used for fast measurement of k_{cat}/K_M values for wild-type and mutant PfuFEN1 enzymes was performed in 10 μl 10 mM Mops (pH 7.5), 50 mM KCl, 7.5 mM $MgCl_2$, 10 ng/μl tRNA including 0.1–5 nM enzyme and 50 nM natural or modified substrate. The reactions were assembled on ice and then incubated at 51 °C for periods of time sufficient to reach 5–20% substrate cleavage. The k_{cat}/K_M values were calculated as the initial cleavage rate divided by the enzyme and initial substrate concentrations under the assumption that the initial substrate concentration was much lower than K_M.

Molecular modeling

PfuFEN1 structure coordinates used in MD simulations were obtained from the PDB database, accession number 1B43[23] and coordinates for the magnesium ions were provided by J. A. Tainer (personal communications). To reduce computation time, the C terminus residues 290–340 of PfuFEN1 were omitted. MD simulations were preformed with AMBER 6.0 software.[62,63] A 2 fs time-step was employed and SHAKE was used for bonds involving hydrogen atoms. The generalized Born (GB) solvation model was used with a non-bonded interactions cut-off distance of either 8 Å or 12 Å and 0.2 M monovalent salt concentration.[64–66] A 1 ns MD required ~six days of computation on a cluster of four computers with dual 1.5 GHz Athlon (AMD) processors and Linux Mandrake 8.1 operating system. A *B*-form DNA duplex of the overlap flap substrate (Figure 2(B)) was generated using the Biopolymer software (Insight II-Accelrys, Inc.) subjected to 100 steps of gradient EM followed by 250 ps of dynamics. To assemble the initial complex, the substrate was manually placed in the active-site DNA-binding groove of PfuFEN1 with an average distance of 3–15 Å from the enzyme with the cleavable phosphodiester linkage facing the M-1 and M-2 metal ions. The MD protocol used here included EM, simulated annealing (SA), and production stages. During the EM stage, the initial PfuFEN1/DNA complex assembly was subjected to 200 steps of conjugate gradient EM to relax energetically unfavorable interactions. Following the EM stage, an SA stage was performed during which the system was gradually heated from 0 K to 500 K for 6 ps using a heat bath time coupling constant of 2 ps, then the system was slowly cooled to 300 K over a period of 14 ps while gradually decreasing the heat bath constant to 0.5 ps. During the SA stage, all enzyme residues except the two magnesium ions, the helical arch (residues 75–127), the β_A/β_B loop (184–210), and the HhH motif (223–256) were constrained to their crystallographic positions using a harmonic potential force constant k_c of 5 kcal/mol/Å^2. Additionally, DNA residues except A_u17–T_u19, T_d1–A_d7, and T_t17–T_t21 were restrained using a flat-well harmonic potential of 2.25(±0.50) Å for Watson–Crick hydrogen bonds and $\alpha = -65(\pm 60)°$, $\beta = -150(\pm 45)°$, $\gamma = 40(\pm 45)°$, $\delta = 120(\pm 90)°$, $\epsilon = 170(\pm 60)°$, $\zeta = -70(\pm 90)°$, and $\chi = -125(\pm 90)°$ for dihedral angels. Finally, flat-well harmonic minimal restraints were used to define a distance of 3.50(±1.50) Å between the M-1 ion and the cleavable phosphodiester linkage P_d6 and a distance of 5.0(±2.50) Å between the magnesium ions, M-1 and M-2, and the active site residues (Table 2). During the SA stage, the restraint penalties were gradually introduced by increasing their harmonic potential force constant k_r from 0 to 15 kcal/mol per Å^2 over the 6 ps heating steps and then to 20 kcal/mol per Å^2 over the 14 ps cooling steps to allow for slow relaxation and prevent abnormal changes in the enzyme or DNA structure. The production stage with minimal restraints was performed using a temperature coupling constant of 1.0 ps, k_c of 1 kcal/mol per Å^2 and k_r of 20 kcal/mol per Å^2 at 300 K for 1 ns. "Residue-specific" restraints included the minimal restraints and an additional set of distance restraints defining specific contacts between the substrate and the enzyme (see Results). Using the final structure of the 1 ns minimal restraints trajectory as starting coordinates, MD simulations with residue-specific restraints were performed with 20 ps SA followed by 1 ns production stage (see above) and 4 ns of additional simulation during which the enzyme constraints were removed for all residues. The final structure was subjected to 500 steps of conjugate gradient EM. Cutoff distances of either 8 Å or 12 Å for non-bonded interactions were used for two initial 1 ns production dynamics with residue-specific restraints. Both cut-off distances produced similar structures with RMSD for heavy atoms of ~1.2 Å. To speed up computation, the 8 Å cut-off distance was used for the 4 ns production dynamics with residue-specific restraints. The average heavy atoms RMSD between consecutive 1 ps structures of 1 ns MD simulations converged to <1.5 Å RMSD after 400–500 ps.

Acknowledgements

We thank James E. Dahlberg, Olke C. Uhlenbeck, and Laura Heisler for helpful suggestions and critical reading of the manuscript. We also thank John A. Tainer and David J. Hosfield for providing the metal coordinates for PfuFEN1.

References

1. Hollingsworth, H. C. & Nossal, N. G. (1991). Bacteriophage T4 encodes an RNase H which removes RNA primers made by the T4 DNA replication system *in vitro*. *J. Biol. Chem.* **266**, 1888–1897.
2. Sayers, J. R. & Eckstein, F. (1990). Properties of overexpressed phage T5 D15 exonuclease. Similarities with *Escherichia coli* DNA polymerase I 5′–3′ exonuclease. *J. Biol. Chem.* **265**, 18311–18317.
3. Deutscher, M. P. & Kornberg, A. (1969). Enzymatic synthesis of deoxyribonucleic acid. XXIX. Hydrolysis of deoxyribonucleic acid from the 5′ terminus by an exonuclease function of deoxyribonucleic acid polymerase. *J. Biol. Chem.* **244**, 3029–3037.
4. Klenow, H. & Overgaard-Hansen, K. (1970). Proteolytic cleavage of DNA polymerase from *Escherichia coli* B into an exonuclease unit and a polymerase unit. *FEBS Letters*, **6**, 25–27.
5. Lyamichev, V., Brow, M. A. & Dahlberg, J. E. (1993).

Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778–783.
6. Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A. & Shen, B. (1998). Newly discovered archaebacterial flap endonucleases show a structure-specific mechanism for DNA substrate binding and catalysis resembling human flap endonuclease-1. *J. Biol. Chem.* **273**, 27154–27161.
7. Hwang, K. Y., Baek, K., Kim, H. Y. & Cho, Y. (1998). The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nature Struct. Biol.* **5**, 707–713.
8. Kaiser, M. W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L. & Lyamichev, V. I. (1999). A comparison of eubacterial and archaeal structure-specific 5′-exonucleases. *J. Biol. Chem.* **274**, 21387–21394.
9. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. (1993). Yeast excision repair gene *RAD2* encodes a single-stranded DNA endonuclease. *Nature*, **366**, 365–368.
10. Harrington, J. J. & Lieber, M. R. (1994). Functional domains within FEN-1 and RAD2 define a family of structure-specific endonucleases: implications for nucleotide excision repair. *Genes Dev.* **8**, 1344–1355.
11. Lindahl, T., Gally, J. A. & Edelman, G. M. (1969). Deoxyribonuclease IV a new exonuclease from mammalian tissues. *Proc. Natl Acad. Sci. USA*, **62**, 597–603.
12. Harrington, J. J. & Lieber, M. R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. *EMBO J.* **13**, 1235–1246.
13. Murante, R. S., Huang, L., Turchi, J. J. & Bambara, R. A. (1994). The calf 5′- to 3′-exonuclease is also an endonuclease with both activities dependent on primers annealed upstream of the point of cleavage. *J. Biol. Chem.* **269**, 1191–1196.
14. O'Donovan, A., Scherly, D., Clarkson, S. G. & Wood, R. D. (1994). Isolation of active recombinant XPG protein, a human DNA repair endonuclease. *J. Biol. Chem.* **269**, 15965–15968.
15. Lundquist, R. C. & Olivera, B. M. (1982). Transient generation of displaced single-stranded DNA during nick translation. *Cell*, **31**, 53–60.
16. Goulian, M., Richards, S. H., Heard, C. J. & Bigsby, B. M. (1990). Discontinuous DNA synthesis by purified mammalian proteins. *J. Biol. Chem.* **265**, 18461–18471. Published erratum appears in *J. Biol. Chem*, **265**, 1990, 22569.
17. Turchi, J. J. & Bambara, R. A. (1993). Completion of mammalian lagging strand DNA replication using purified proteins. *J. Biol. Chem.* **268**, 15136–15141.
18. Murray, J. M., Tavassoli, M., al-Harithy, R., Sheldrick, K. S., Lehmann, A. R., Carr, A. M. & Watts, F. Z. (1994). Structural and functional conservation of the human homolog of the *Schizosaccharomyces pombe rad2* gene, which is required for chromosome segregation and recovery from DNA damage. *Mol. Cell Biol.* **14**, 4878–4888.
19. O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. (1994). XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. *Nature*, **371**, 432–435.
20. Reynaldo, L. P., Vologodskii, A. V., Neri, B. P. & Lyamichev, V. I. (2000). The kinetics of oligonucleotide replacements. *J. Mol. Biol.* **297**, 511–520.
21. Lyamichev, V., Brow, M. A., Varvel, V. E. & Dahlberg, J. E. (1999). Comparison of the 5′ nuclease activities of *Taq* DNA polymerase and its isolated nuclease domain. *Proc. Natl Acad. Sci. USA*, **96**, 6143–6148.
22. Kao, H. I., Henricksen, L. A., Liu, Y. & Bambara, R. A. (2002). Cleavage specificity of *Saccharomyces cerevisiae* flap endonuclease 1 suggests a double-flap structure as the cellular substrate. *J. Biol. Chem.* **277**, 14379–14389.
23. Hosfield, D. J., Mol, C. D., Shen, B. & Tainer, J. A. (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, **95**, 135–146.
24. Matsui, E., Musti, K. V., Abe, J., Yamasaki, K., Matsui, I. & Harata, K. (2002). Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1. *J. Biol. Chem.* **277**, 37840–37847.
25. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612–616.
26. Ceska, T. A., Sayers, J. R., Stier, G. & Suck, D. (1996). A helical arch allowing single-stranded DNA to thread through T5 5′-exonuclease. *Nature*, **382**, 90–93.
27. Mueser, T. C., Nossal, N. G. & Hyde, C. C. (1996). Structure of bacteriophage T4 RNase H, a 5′ to 3′ RNA–DNA and DNA–DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. *Cell*, **85**, 1101–1112.
28. Shen, B., Nolan, J., Sklar, L. & Park, M. (1996). Essential amino acids for substrate binding and catalysis of human flap endonuclease 1. *J. Biol. Chem.* **271**, 9173–9176.
29. Murante, R. S., Rust, L. & Bambara, R. A. (1995). Calf 5′ to 3′ exo/endonuclease must slide from a 5′ end of the substrate to perform structure-specific cleavage. *J. Biol. Chem.* **270**, 30377–30383.
30. Bornarth, C. J., Ranalli, T. A., Henricksen, L. A., Wahl, A. F. & Bambara, R. A. (1999). Effect of flap modifications on human FEN1 cleavage. *Biochemistry*, **38**, 13347–13354.
31. Doherty, A. J., Serpell, L. C. & Ponting, C. P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucl. Acids Res.* **24**, 2488–2497.
32. Shao, X. & Grishin, N. V. (2000). Common fold in helix-hairpin-helix proteins. *Nucl. Acids Res.* **28**, 2643–2650.
33. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108–4120.
34. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H. & Kraut, J. (1996). Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry*, **35**, 12742–12761.
35. Shen, B., Nolan, J. P., Sklar, L. A. & Park, M. S. (1997). Functional analysis of point mutations in human flap endonuclease-1 active site. *Nucl. Acids Res.* **25**, 3332–3338.
36. Bhagwat, M., Meara, D. & Nossal, N. G. (1997). Identification of residues of T4 RNase H required for catalysis and DNA binding. *J. Biol. Chem.* **272**, 28531–28538.
37. Xu, Y., Derbyshire, V., Ng, K., Sun, X. C., Grindley, N. D. & Joyce, C. M. (1997). Biochemical and mutational studies of the 5′–3′ exonuclease of DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.* **268**, 284–302.
38. Garforth, S. J., Ceska, T. A., Suck, D. & Sayers, J. R. (1999). Mutagenesis of conserved lysine residues in bacteriophage T5 5′–3′ exonuclease suggests separate mechanisms of endo and exonucleolytic cleavage.

Proc. Natl Acad. Sci. USA (in process Citation), **96**, 38–43.
39. Qiu, J., Bimston, D. N., Partikian, A. & Shen, B. (2002). Arginine residues 47 and 70 of human flap endonuclease-1 are involved in DNA substrate interactions and cleavage site determination. *J. Biol. Chem.* **277**, 24659–24666.
40. Dervan, J. J., Feng, M., Patel, D., Grasby, J. A., Artymiuk, P. J., Ceska, T. A. & Sayers, J. R. (2002). Interactions of mutant and wild-type flap endonucleases with oligonucleotide substrates suggest an alternative model of DNA binding. *Proc. Natl Acad. Sci. USA*, **99**, 8542–8547.
41. Dertinger, D. & Uhlenbeck, O. C. (2001). Evaluation of methylphosphonates as analogs for detecting phosphate contacts in RNA–protein complexes. *RNA*, **7**, 622–631.
42. Noble, S. A., Fisher, E. F. & Caruthers, M. H. (1984). Methylphosphonates as probes of protein–nucleic acid interactions. *Nucl. Acids Res.* **12**, 3387–3404.
43. Botfield, M. C. & Weiss, M. A. (1994). Bipartite DNA recognition by the human Oct-2 POU domain: POUs-specific phosphate contacts are analogous to those of bacteriophage lambda repressor. *Biochemistry*, **33**, 2349–2355.
44. Smith, S. A. & McLaughlin, L. W. (1997). Probing contacts to the DNA backbone in the trp repressor–operator sequence-specific protein–nucleic acid complex using diastereomeric methylphosphonate analogues. *Biochemistry*, **36**, 6046–6058.
45. Hou, Y. M., Zhang, X., Holland, J. A. & Davis, D. R. (2001). An important 2′-OH group for an RNA–protein interaction. *Nucl. Acids Res.* **29**, 976–985.
46. Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. & Gait, M. J. (1994). Methylphosphonate mapping of phosphate contacts critical for RNA recognition by the human immunodeficiency virus tat and rev proteins. *Nucl. Acids Res.* **22**, 2592–2600.
47. Tomky, L. A., Strauss-Soukup, J. K., Maher, L. J. & 3rd, . (1998). Effects of phosphate neutralization on the shape of the AP-1 transcription factor binding site in duplex DNA. *Nucl. Acids Res.* **26**, 2298–29305.
48. Goodsell, D. S., Kopka, M. L., Cascio, D. & Dickerson, R. E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl Acad. Sci. USA*, **90**, 2930–2934.
49. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*, W. H. Freeman and company, New York pp. 1036–1037.
50. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. (1997). Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*, **36**, 11205–11215.
51. Hollis, T., Ichikawa, Y. & Ellenberger, T. (2000). DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli AlkA*. *EMBO J.* **19**, 758–766.
52. Xu, Y., Potapova, O., Leschziner, A. E., Grindley, N. D. & Joyce, C. M. (2001). Contacts between the 5′ nuclease of DNA polymerase I and its DNA substrate. *J. Biol. Chem.* **276**, 30167–30177.
53. Ceska, T. A. & Sayers, J. R. (1998). Structure-specific DNA cleavage by 5′ nucleases. *Trends Biochem. Sci.* **23**, 331–336.
54. Beese, L. S., Derbyshire, V. & Steitz, T. A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science*, **260**, 352–355.
55. Katahira, M., Miyanoiri, Y., Enokizono, Y., Matsuda, G., Nagata, T., Ishikawa, F. & Uesugi, S. (2001). Structure of the C-terminal RNA-binding domain of hnRNP D0 (AUF1), its interactions with RNA and DNA, and change in backbone dynamics upon complex formation with DNA. *J. Mol. Biol.* **311**, 973–988.
56. Dertinger, D., Dale, T. & Uhlenbeck, O. C. (2001). Modifying the specificity of an RNA backbone contact. *J. Mol. Biol.* **314**, 649–654.
57. Kim, C. Y., Park, M. S. & Dyer, R. B. (2001). Human flap endonuclease-1: conformational change upon binding to the flap DNA substrate and location of the Mg^{2+} binding site. *Biochemistry*, **40**, 3208–3214.
58. McGaughey, G. B., Gagne, M. & Rappe, A. K. (1998). pi-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **273**, 15458–15463.
59. Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G. A., Frouin, I. *et al.* (2001). Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. *Proc. Natl Acad. Sci. USA*, **98**, 14298–14303.
60. Bae, S. H., Bae, K. H., Kim, J. A. & Seo, Y. S. (2001). RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature*, **412**, 456–461.
61. Gray, D. M., Hung, S. H. & Johnson, K. H. (1995). Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.* **246**, 19–34.
62. Case, D. A., Pearlman, D. A., Caldwell, J. C. III, Cheatham, T. E., Ross, W. S., Simmerling, C. L., *et al.* (1999). *AMBER 6*, Univerity of California, San Francisco.
63. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G. *et al.* (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
64. Tsui, V. & Case, D. A. (2000). Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.
65. Onufriev, A., Bashford, D. & Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *J. Phys. Chem.* **104**, 3712–3720.
66. Tsui, V. & Case, D. A. (2000). Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **122**, 2489–2498.

Edited by K. Morikawa

(Received 3 January 2003; received in revised form 25 February 2003; accepted 3 March 2003)